

The final publication is available at link.springer.com via <http://dx.doi.org/10.1007/s11042-017-5276-7>

Review on Computer Vision Techniques in Emergency Situations

Laura Lopez-Fuentes · Joost van de Weijer · Manuel González-Hidalgo · Harald Skinnemoen · Andrew D. Bagdanov

Received: date / Accepted: date

Abstract In emergency situations, actions that save lives and limit the impact of hazards are crucial. In order to act, situational awareness is needed to decide what to do. Geolocalized photos and video of the situations as they evolve can be crucial in better understanding them and making decisions faster. Cameras are almost everywhere these days, either in terms of smartphones, installed CCTV cameras, UAVs or others. However, this poses challenges in big data and information overflow. Moreover, most of the time there are no disasters at any given location, so humans aiming to detect sudden situations may not be as alert as needed at any point in time. Consequently, computer vision tools can be

This work was partially supported by the Spanish Grants TIN2016-75404-P AEI/FEDER, UE, TIN2014-52072-P, TIN2013-42795-P and the European Commission H2020 I-REACT project no. 700256. Laura Lopez-Fuentes benefits from the NAERINGSPHD fellowship of the Norwegian Research Council under the collaboration agreement Ref.3114 with the UIB. We thank the NVIDIA Corporation for support in the form of GPU hardware.

L. Lopez-Fuentes
TM-RCS Working Group, AnsuR Technologies AS, Fornebu, Norway.
University of the Balearic Islands, Dept of Mathematics and Computer Science, Crta. Valldemossa, km 7.5, E-07122 Palma, Spain.
Computer Vision Center. Universitat Autònoma de Barcelona. 08193-Barcelona. Spain. E-mail: l.lopez@uib.es

J. van de Weijer
Computer Vision Center. Universitat Autònoma de Barcelona. 08193-Barcelona. Spain. E-mail: joost@cvc.uab.es

M. González-Hidalgo
University of the Balearic Islands, Dept of Mathematics and Computer Science, Crta. Valldemossa, km 7.5, E-07122 Palma, Spain. E-mail: manuel.gonzalez@uib.es

H. Skinnemoen
TM-RCS Working Group, AnsuR Technologies AS, Fornebu, Norway.
E-mail: harald@ansur.no

A. D. Bagdanov
University of Florence, DINFO, Via Santa Marta 3, 50139 Florence, Italy. E-mail: andrew.bagdanov@unifi.it

an excellent decision support. The number of emergencies where computer vision tools has been considered or used is very wide, and there is a great overlap across related emergency research. Researchers tend to focus on state-of-the-art systems that cover the same emergency as they are studying, obviating important research in other fields. In order to unveil this overlap, the survey is divided along four main axes: the types of emergencies that have been studied in computer vision, the objective that the algorithms can address, the type of hardware needed and the algorithms used. Therefore, this review provides a broad overview of the progress of computer vision covering all sorts of emergencies.

Keywords Emergency management · computer vision · decision makers · situational awareness · critical situation

1 Introduction

Emergencies are a major cause of both human and economic loss. They vary in scale and magnitude, from small traffic accidents involving few people, to full-scale natural disasters which can devastate countries and harm thousands of people. *Emergency management* has become very important to reduce the impact of emergencies. As a consequence, using modern technology to implement innovative solutions to prevent, mitigate, and study emergencies is an active field of research.

The presence of digital cameras has grown explosively over the last two decades. Camera systems perform real-time recording of visual information in hotspots where an accident or emergency is likely to occur. Examples of such systems are cameras on the road, national parks, inside banks, shops, airports, metros, streets and swimming pools. In Figure 1 we give some examples of frames from videos that could potentially be recorded by some of these monitoring systems. It is estimated that as of 2014 there were over 245

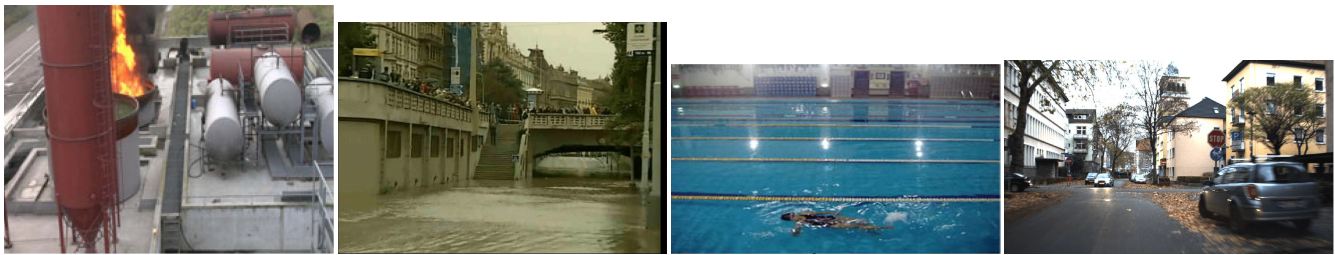


Fig. 1: Example frames from videos that could potentially be recorded to detect different emergencies: a factory being monitored to for rapid fire detection [156], a street monitored to detect floods [16], a surveillance camera on a pool to detect drowning persons, [26] and a vehicle with on board camera for driver assistance [69].

million active, operational and professionally installed video cameras around the world [76]. Mounting these video cameras has become very cheap but there are insufficient human resources to observe their output. Therefore, most of the data is being processed *after* an emergency has already occurred, thus losing the benefit of having a real-time monitoring tool in the first place.

The problem of real-time processing of the visual output from monitoring systems in potential emergency scenarios has become a hot topic in the field of computer vision. As we will show in this review, computer vision techniques have been used to assist in various stages of emergency management: from prevention, detection, assistance of the response, to the understanding of emergencies. A myriad of algorithms has been proposed to prevent events that could evolve into disaster and emergency situations. Others have focused on fast response after an emergency has already occurred. Yet other algorithms focus on assisting during the response to an emergency situation. Finally, several methods focus on improving our understanding of emergencies with the aim of predicting the likelihood of an emergency or dangerous situation occurring.

This review provides a broad overview of the progress of computer vision covering all sorts of emergencies. The review is organized in four main axes along which researchers working on the topic of computer vision for emergency situations have organized their investigations. The first axis relates to the *type of emergency* being addressed, and the most important division along this axis is the distinction between *natural* and *man-made* emergencies. The second axis determines the objective of a system: if it is going to detect risk factors to prevent the emergency, help detecting the emergency once it has occurred, provide information to help emergency responders to react to the emergency, or model the emergency to extract valuable information for the future. The third axis determines where the visual sensors should be placed, and also what type of sensors are most useful for the task. The last axis determines which algorithms will be

used to assist in analyzing the data. This typically involves a choice of computer vision algorithms combined with machine learning techniques.

When organizing emergency situations along these axes, our review shows that there is significant overlap across emergency research areas, especially in the sensory data and algorithms used. For this reason this review is useful for researchers in emergency topics, allowing them to more quickly find relevant works. Furthermore, due to its general nature, this article aims to increase the flow of ideas among the various emergency research fields.

Since it is such an active field of research, there are several reviews on computer vision algorithms in emergency situations. However these reviews focus on a single, specific topic. In the field of fire detection, a recent survey was published summarizing the different computer vision algorithms proposed in the literature to detect fire in video [25]. Also Gade et al. [50] published a survey on applications of thermal cameras in which fire detection is also considered. Another related research direction is human activity recognition and detection of abnormal human actions. This field has many applications to video surveillance, as shown in a recent review on this topic [86]. Work in the area of human activity recognition has focused on the specific topic of falling person detection. Mubashir et al. [50] describe the three main approaches to falling person detection, one of them being computer vision. Another important application area of emergency management is traffic. The number of vehicles on roads has significantly increased in recent years, and simultaneously the number of traffic accidents has also increased. Therefore a large number of road safety infrastructure and control systems have been developed. Three main reviews have been recently published on the topic of road vehicles monitoring and accident detection using computer vision [20, 85, 92]. As discussed here, there are several review papers about different specific emergency situations, however there is *no one* work reviewing computer vision in emergency situations considering many types of emergency

situations. By doing so, we aspire to draw a wider picture of the field, and to promote cross-emergency dissemination of research ideas.

Computer vision is just one of several scientific fields which aims to improve emergency management. Before going into depth on computer vision techniques for emergency management, we briefly provide some pointers to other research fields for interested readers. Emergency management is a hot topic in the field of Internet of Things (IoT) [108, 171]. The goal of the IoT is to create an environment in which information from any object connected to the network be efficiently shared with others in real-time. The authors of [171] provide an excellent study on how IoT technology can enhance emergence rescue operations. Robotics research is another field which already is changing emergency management. Robots are being studied for handling hazardous situations with radiation [120], for assisting crowd evacuation [133], and for search and rescue operations [136]. Finally, we mention reviews on human centered sensing for emergency evacuation [126] and research on the impact of social media on emergency management. The main advantage of computer vision compared to these other research fields is the omni-presence of cameras in modern society, both in the public space and mobile phones of users. However, for some problems computer vision might currently be still less accurate than other techniques.

2 Organization of this Review

Before we enter into the organization of the review it is important to define what we consider as an *emergency* and, therefore, what situations and events will be encompassed in this definition. We consider as an emergency any sudden or unexpected situation that meets at least one of the following conditions:

- it is a situation that represents an imminent risk of health or life to a person or group of persons or damage to properties or environment;
- it is a situation that has already caused loss of health or life to a person or group of persons or damage to properties or environment; or
- it is a situation with high probability of scaling, putting in risk the health or life of a person or group of persons or damage to properties or environment.

An emergency situation, therefore, requires situational awareness, rapid attention, and remedial action to prevent degradation of the situation. In some cases emergency situations can be prevented by avoiding and minimizing risks. Also, in some situations, the occurrence of an emergency situation can help to understand risk factors and improve the prevention mechanisms. Taking this definition in consideration, any type of uncontrolled fire would be considered

as an emergency situation as it is an unexpected situation that could represent a risk for persons, the environment, and property. A controlled demolition of a building would not be considered an emergency situation, as by definition if it is controlled it is not an unexpected situation any more. However, if someone is injured or something unforeseen was damaged during the demolition, that would then be considered an emergency situation.

After having defined what we consider an emergency situation we briefly explain the organization of this review. We begin in Section 3 with an overview and classification into subcategories of emergency situations. Here, we briefly introduce the three axes along which we will organize the existing literature on computer vision for emergency situations, which corresponds to Sections 4 to 6. The organization is also summarized in Figure 2.

The Monitoring Objective Axis (Section 4): the management of an emergency has a life cycle which can be divided into four main phases [121]:

- **Preparedness:** during this phase there are risk factors that could trigger an emergency at some point. In order to reduce the chance of an emergency happening, it is important to estimate these factors and avoid or minimize them as much as possible. Not all emergencies can be prevented, but in some cases preventive measures can help to reduce the impact of a disaster (like teaching people how to behave during a hurricane).
- **Response:** response activities are performed after the emergency has occurred. At this point, it is very important to detect and localize the emergency and to maintain situational awareness to have a rapid and efficient reaction. Often the minutes immediately after the occurrence of an emergency can be crucial to mitigating its effects.
- **Recovery:** recovery activities take place until normality is recovered. In this phase actions must be taken to save lives, prevent further damaging or escalation of the situation. We consider that normality has been recovered when the immediate threat to health, life, properties or environment has subsided.
- **Mitigation:** this phase starts when normality has been re-established. During this phase it is important to study and understand the emergency, how it could have been avoided or what responses could have helped to mitigate it.

The Sensors and Acquisition System Axis (Section 5):

An emergency has four main management phases: prevention, detection, response and understanding of an emergency. For a computer vision algorithm to help in one or more of these phases, it is important to have visual sensors to record the needed information. There are many different types of devices for this, as also many types of devices in which they may be installed depending on the type of data being

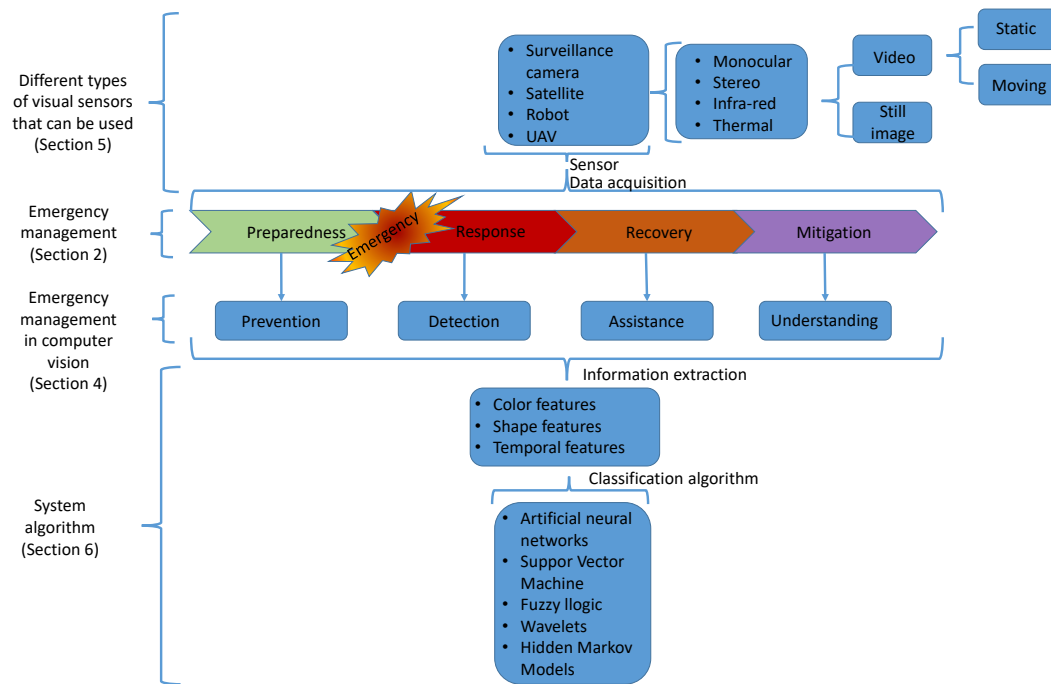


Fig. 2: A graphical overview of the organization of this survey.

| Generic Emergency Group | Emergency sub-group | Emergency Type |
|-------------------------|---------------------|---------------------|
| Natural emergencies | | Fire |
| | | Flood |
| | | Drought |
| | | Earthquake |
| | | Hurricane/Tornado |
| Man-made emergencies | | Landslide/Avalanche |
| | Single person | Falling person |
| | | Drowning |
| | | Injured civilians |
| | Multiple person | Road accident |
| | | Crowd related |
| Weapon threaten | | |

Table 1: Types of emergencies and their classification.

captured. Some of the most common are: fixed monitoring cameras, satellites, robots and Unmanned Aerial Vehicles (UAVs). These devices can capture different types of visual data, such as stereo or monocular images, infrared or thermal. Depending on the rate of data acquisition this data can be in the form of video or still images, and the video can be either from a static or moving source.

The Algorithmic Axis (Section 6): The visual data collected by the sensor is then subject to a feature extraction process. Finally the features extracted are used as an input to a modeling or classification algorithm. In Figure 2 the main feature extraction and classification algorithms are listed.

3 Types of Emergencies

There are many types of emergencies and a great deal of effort has been invested in classifying and defining them [1, 12]. However there are many emergencies that due to their nature have not been studied in computer vision, such as famine or pollution. These types of emergencies do not have clear visual consequences that can be easily detected by visual sensors. In this paper we focus only on emergencies relevant in the computer vision field. In Table 1 we classify the types of emergencies that we consider in this survey. We distinguish two generic emergency groups: natural emergencies and man-made emergencies. Natural emergencies are divided into three emergency types: fire, flood and drought.

Emergencies caused by humans are divided into two subtypes depending on the scope: emergencies that cause dangers to multiple persons and emergencies that cause dangers to a single person.

3.1 Natural emergencies

Natural emergencies cover any adverse event resulting from natural processes of the earth. Natural emergencies can result in a loss of lives and can have huge economic costs. The severity of the emergency can be measured by the losses and the ability of the population to recuperate. Although natural emergencies are external to humans and therefore can not be prevented, society has been increasingly able to reduce the impact of these natural events by building systems for early detection or assistance during the event. Among these systems built to detect or assist natural emergencies, computer vision has played an important role in emergencies related to fire, flood and drought.

Fire: electric fire detectors have been integrated into buildings since the late 1980s. These devices detect smoke, which is a typical fire indicator. Currently, two major types of electric smoke detector devices are used [127]: ionization smoke detectors and photoelectric smoke detectors. Although the devices achieve very good performance in closed environments, their reliability decreases in open spaces because they rely on smoke particles arriving to the sensor. For this reason, fire detection in monitoring cameras has become an important area of research in computer vision. The use of video as input in fire detection makes it possible to use in large and open spaces. Hence, most computer vision algorithms that study fire detection focus mainly on open spaces [39,61]. However, some articles [89,153] also study fire detection in closed spaces, claiming to have several advantages over conventional smoke detectors: visual fire detectors can trigger the alarm before conventional electric devices, as they do not rely on the position of fire particles, and they can give information about the position and size of the fire.

Flood: flood results from water escaping its usual boundaries in rivers, lakes, oceans or man-made canals due to an accumulation of rainwater on saturated ground, a rise in the sea level or a tsunami. As a result, the surrounding lands are inundated. Floods can result in loss of life and can also have huge economic impact by damaging agricultural land and residential areas. Flood detection or modeling algorithms are still at a very early stage. In computer vision, there are two main trends: algorithms based on still cameras monitoring flood-prone areas [96] and systems that process satellite images [111,112,113]. On the one hand, algorithms based on earth still cameras cover a limited area but do not have hardware and software restrictions. On the other hand, algorithms based on satellite imagery cover a much wider area,

but if the system is to be installed on board the satellite, there are many factors that have to be considered when deciding the type of hardware and software to be used on board a satellite [90,112,140].

Drought: droughts are significant weather-related disasters that can result in devastating agricultural, health, economic and social consequences. There are two main characteristics that determine the presence of drought in an area: a low surface soil moisture and a decrease in rainfall. Soil moisture data is rarely computed with visual data, however in [66] they use high-resolution remotely sensed data to determine the water balance of the soil. Another technique used to determine the beginning of drought is to compute the difference between the average precipitation, or some other climatic variable, over some time period and comparing the current situation to a historical record. Observation of rainfall levels can help prevent or mitigate drought. Although rainfall does not have a clear visual consequence that can be monitored through cameras, there have been some image processing algorithms studied to map annual rainfall records [49,140]. These algorithms help to determine the extension and severity of a drought and to determine regions likely to suffer from this emergency.

Earthquake: earthquakes are sudden shakes or vibrations of the ground caused by tectonic movement. The magnitude of an earthquake can vary greatly from almost imperceptible earthquakes (of which hundreds are registered every day) up to very violent earthquakes that can be devastating. Earthquake prediction is a whole field of research called seismology. However, due to the lack of distinctive visual information about an earthquake, this emergency has received relatively little attention from the computer vision community. Several works address the robustness of motion tracking techniques under seismic-induced motions [40,73], with the aim of assisting rescue and reconnaissance crews. The main difficulty for these systems is the fact that the camera itself is also undergoing motion. In [40] this problem is addressed by introducing a dynamic camera reference system which is shared between cameras and allows tracking of objects with respect to this reference coordinate system. Also, computer vision plays a role during the rescue of people affected by these natural disaster by detecting people with robots or UAVs in the damaged areas, this is tackled in the “injured civilians” section.

Hurricane/tornado: hurricanes and tornadoes are violent storms characterized by heavy wind and rain. Similarly to earthquakes, depending on their magnitude they can be almost unnoticeable or devastating and at the same time the visual information derived from these natural disasters has not yet been studied in the field of computer vision. However, similarly to earthquakes, robots and UAVs can be used

to find people in the areas affected by the disaster using human detection algorithms.

The more general work on “injured civilians” detection in affected areas which is shared among any other emergency that involves people rescue like earthquakes and hurricanes, is again applicable to landslide and avalanche emergencies. We found one work which evaluates this in particular for avalanches, and which takes into account the particular environment where avalanches occur [11].

3.2 Emergencies caused by humans

Emergencies caused by humans cover adverse events and disasters that result from human activity or man-made hazards. Although natural emergencies are considered to generate a greater number of deaths and have larger economic impact, emergencies caused by humans are increasing in importance and magnitude. These emergencies can be unintentionally caused, such as road accidents or drowning persons, but they can also be intentionally provoked, such as armed conflicts. In this survey, emergencies caused by humans are divided in two types depending on their scope: emergencies affecting multiple persons and emergencies affecting a single person.

3.2.1 Emergencies affecting multiple persons

These emergencies are human-caused and affects several persons.

Road accident emergencies: Although depending on the extent of the emergency they can affect one or multiple persons, we consider road accidents to be multiple-person emergencies. Transport emergencies result from human-made vehicles. They occur when a vehicle collides with another vehicle, a pedestrian, animal or another road object. In recent years, the number of vehicles on the road has increased dramatically and accidents have become one of the highest worldwide death factors, reaching the number one cause of death among those aged between 15 and 29 [159]. In order to prevent these incidents, safety infrastructure has been developed, many monitoring cameras have been installed on roads, and different sensors have been installed in vehicles to prevent dangerous situations. In computer vision there are three main fields of study in which have focused on the detection or prevention of transport emergencies.

- Algorithms which focus on detecting anomalies, as in most cases they correspond to dangerous situations [78, 134, 145].
- Algorithms specialized on detecting a concrete type of road emergency, such as a collision [83, 157, 176].
- Algorithms based on on-board cameras which are designed to prevent dangerous situations, such as algorithms that detect pedestrians and warn the driver [53, 62, 64].

Crowd-related emergencies: These type of emergencies are normally initiated by another emergency when man persons are gathered in a limited area. They normally occur in multitudinous events such as sporting events, concerts or strikes. These type of events pose a great risk because when any emergency strikes it can lead to collective panic, thus aggravating the initial emergency. This is a fairly new area of research in computer vision. In computer vision for crowd-related emergencies we find different sub-topics like crowd density estimation, analysis of abnormal events [6], crowd motion tracking [75] and crowd behaviour analysis [51], among others.

Weapon threat emergencies: Although detecting weapons through their visual features and characteristics has not yet been studied in computer vision, some studies have been done in the direction of detecting the heat produced by them. A dangerous weapon that threatens innocent people every year are mines. Millions of old mines are hidden under the ground of many countries which kill and injure people. These mines are not visible from the ground but they emit heat that can be detected using thermal cameras [119, 138, 164]. Another weapon that can be detected through thermal sensors are guns, due to the heat that they emit when they are fired [124].

3.2.2 Emergencies affecting a single person

These are emergencies are caused by or affect one person. The classic example of this is detecting human falling events for monitoring the well-being of the elderly.

Drowning: Drowning happens when a person lacks air due to an obstruction of the respiratory track. These situations normally occur in swimming pools or beach areas. Although most public swimming pools and beaches are staffed with professional lifeguards, drowning or near-drowning incidents still occur due to staff reaction time. Therefore there is a need to automate the detection of these events. One way of automating detection of these events is the installation of monitoring cameras equipped with computer vision algorithms to detect drowning persons in swimming pools and beach areas. However, drowning events in the sea are difficult to monitor due to the extent of the hazardous area, so most computer vision research on this topic has been focused in drowning people in swimming pools [42, 180].

Injured person: An injured person is a person that has come to some physical harm. After an accident, reaction time is crucial to minimize human losses. Emergency responders focus their work on the rescue of trapped or injured persons in the shortest time possible. A method that would help speed up the process is the use of aerial and ground robots with visual sensors that record the affected area at the same time as detect human victims. Also, in the event

of an emergency situation, knowing if there are human victims in the area affected by the event may help determine the risk the responders are willing to take. For this purpose autonomous robots have been built to automatically explore the area and detect human victims without putting at risk the rescue staff [7, 22, 88, 99, 131, 142]. Detecting injured or trapped humans in a disaster area is an extremely challenging problem from the computer vision perspective due to the articulated nature of the human body and the uncontrolled environment.

Falling person: With the growth of the elderly population there is a growing need to create timely and reliable smart systems to help and assist seniors living alone. Falls are one of the major risks the elderly living alone, and early detection of falls is very important. The most common solution used today are call buttons, which when the senior can press to directly call for help. However these devices are only useful if the senior is conscious after the fall. This problem has been tackled from many different perspectives. On the one hand wearable devices based on accelerometers have been studied and developed [28, 182], some of these devices are in process of being commercialized [44]. On the other hand, falling detection algorithms have been studied from a computer vision perspective [101, 129].

4 Monitoring Objective

As stated in Section 2, an emergency has four main phases which at the same time lead to four emergency management phases: prevention, detection, response/assistance and understanding of emergencies. It is important to understand, study and identify the different phases of an emergency life cycle since the way to react to it changes consequently. In this section, we will classify computer vision algorithms for emergency situations depending on the emergency management phase in which they are used.

4.1 Emergency Prevention

This phase focuses on preventing hazards and risk factors to minimize the likelihood that an emergency will occur. The study of emergency prevention on computer vision can not be applied to any emergency since not all of them are induced by risk factors that have relevant visual features that can be captured by optical sensors. For example, the only visual feature which is normally present before a fire starts is smoke. Therefore research on fire prevention in computer vision focuses on smoke detection. As an example, in Figure 3 (a) we see a region of forest with a dense smoke that looks like the starting of a fire, however a fire detection algorithm would not detect anything until the fire is present in

the video. Some early work on this topic started at the beginning of the century [29, 56]. Since then, smoke detection for fire prevention or early fire detection has been an open topic in computer vision. Most of the work done on smoke detection is treated as a separate problem from fire detection and aims to detect smoke at the smoldering phase to prevent fire [60, 166, 172, 175]. However there is also some work on smoke and flame detection when the fire has already started since smoke in many cases can be visible from a further distance [91, 174].

Traffic accidents are considered as one of the main sources of death and injuries globally. Traffic accidents can be caused by three main risk factors: the environment, the vehicle and the driver. Among these risk factors, it is estimated that the driver accounts for around 90% of road accidents [154]. To avoid these type of accidents, a lot of security devices have been incorporated on vehicles in the past years. There are two types of security devices: “active” whose function is to decrease the risk of an accident occurring, and “passive” which try to minimize the injuries and deaths in an accident. Among the active security devices, there is work on the usage of visual sensors to detect situations that could lead to an accident and force actions that could be taken to prevent the accident [152].

There are many different kind of traffic accidents but one of the most common and at the same time more deadly ones are pedestrians being run over. Due to the variant and deformable nature of humans, the environment variability, the reliability, speed and robustness that a security system is required, the detection of unavoidable impact with pedestrians has become a challenge in computer vision. This field of study is well defined and with established benchmarks and it has served as a common topic in numerous review articles [13, 41, 54, 165]. Now, with the increase of training data and computing power this investigation has had a huge peak in the past years and even the first devices using this technology have already been commercialized [117, 146]. Figure 3 (b) shows an example of the output from a pedestrian detection algorithm on board a vehicle.

Another contributing factor to many accidents is the drivers fatigue, distraction or drowsiness. A good solution to detect the state of the driver in a non intrusive way is using an on board camera able to detect the drivers level of attention to the road. There are different factors that can be taken into account to determine the level of attention to the road of a driver. In [31] they determine if the driver is looking at the road through the head pose and gaze direction, in Figure 3 (c) we show some examples of the images they used to train their algorithm. While in [64] they determine 17 feature points from the face that they consider relevant to detect the level of drowsiness of a person. In [70] they study the movements of the eyelid to detect sleepy drivers.

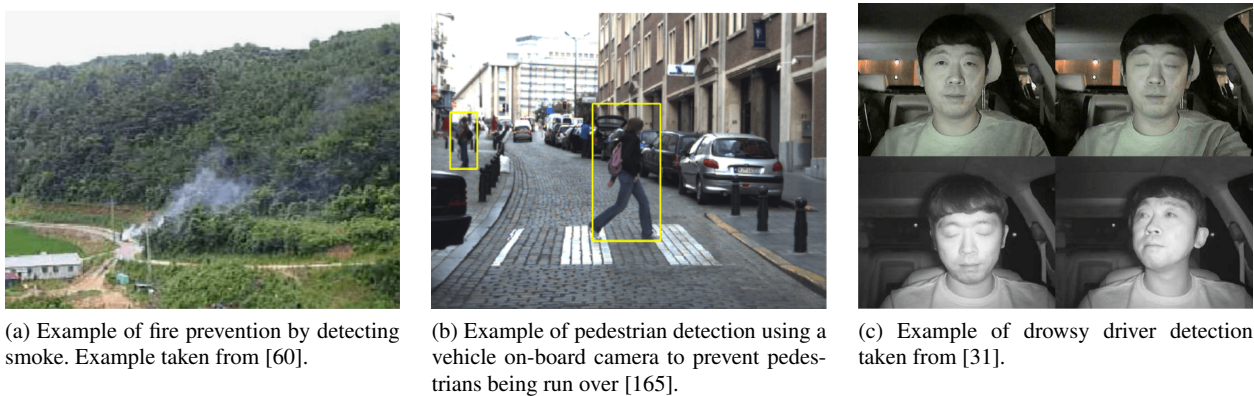


Fig. 3: Example situations in which computer vision techniques have been applied for emergency prevention.

Although not so widely studied, there are other type of road accidents that could also be prevented with the help of computer vision. For example, in [67] they propose a framework to predict unusual driving behavior which they apply to right turns at intersections. Algorithms that use on board cameras to detect the road have also been studied [62], this could help to detect if a vehicle is about to slip on the road. In [9] they propose a system with on board stereo cameras that assist drivers to prevent collisions by detecting sudden accelerations and self-occlusions. Multiple algorithms for traffic sign detection with on board cameras have also widely proposed in the literature [69, 118], these algorithms mostly go in the direction of creating autonomous vehicles, however they can also be used to detect if the driver is not following some traffic rules and is therefore increasing the risk of provoking an accident.

In order to prevent falls, in [101] they propose a method to detect slips arguing that a slip is likely to lead to a fall or that a person that just had a slip is likely to have balance issues that increase the risk of falling down in the near future. They consider a slip as a sequence of unbalanced postures.

4.2 Emergency Detection

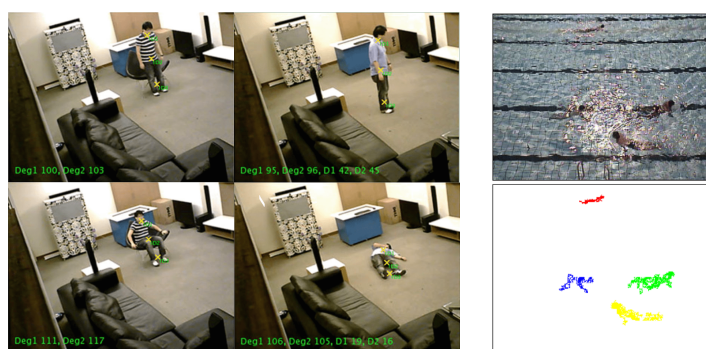
The field of emergency detection through computer vision in contrast with the field of prevention is much wider and has been studied in more depth. In this section we will present the visual information that is being studied to detect different emergencies.

Fire detection has been widely studied in computer vision since it is very useful on open spaces, where other fire detection algorithms fail. In some cases it is useful to not only detect the fire but also some properties of the fire such as the flame height, angle or width [39]. In order to detect fire on visual content, the algorithms studied exploit visual and temporal features of fire. In early research, flame was

the main visual feature that was studied in fire detection [24, 61, 89, 128, 153, 156] however, as stated in Section 4.1, recently smoke has also been studied in the field of fire detection mainly because it appears before fire and it spreads faster so in most cases smoke will appear faster in the camera field of view [91, 174]. As fire has a non-rigid shape with no well defined edges, the visual content used to detect fire is related to the characteristic color of fire. This makes the algorithms in general vulnerable to lighting conditions and camera quality.

Water, the main cue for flood detection, has similar visual difficulties as fire in order to be detected since it has a non-rigid shape and the edges are also not well defined. Moreover, water does not have a specific characteristic color because it might be dirty and it also reflects the surrounding colors. Due to the mentioned difficulties flood detection in computer vision has not been so widely studied. In [96] they use color information and background detail (because as said before the background will have a strong effect on the color of the water) and the ripple pattern of water. In [16] they analyse the rippling effect of water and they use the relationship between the averages for each color channel as a detection metric.

Detecting person falls has become a major interest in computer vision, specially motivated to develop systems for elderly care monitoring at home. In order to detect a falling person it is important to first detect the persons which are visible in the data and then analyse the movements. Most algorithms use background subtraction to detect the person since in indoor situations, most of the movement come from persons. Then, to determine if the movement of the person corresponds to a fall some algorithms use silhouette or shape variation [101, 116, 130], direction of the main movement [68], or relative height with respect to the ground [118]. However, sometimes it is difficult to distinguish between a person falling down and a person sitting, lying or sleeping,



(a) Two very different scenarios such as sitting and falling may be visually similar, in order to differentiate them some researchers suggest to use temporal information [33].

(b) Example of a person detection algorithm in a swimming pool [42].

Fig. 4: Example situations in which computer vision techniques have been applied for emergency detection.

to address this issue, most systems consider the duration of the event since a fall is usually a faster movement [102, 118], for example, in [33] they consider 10 frames as a fast movement and use this threshold to differentiate two visually similar movements such as the ones given in Figure 4 (a) which corresponds to sitting and falling.

The problem of drowning person detection is fairly similar to the falling person detection, since it is also in a closed environment and related to human motion. Similarly to fall detection algorithms, the first step to detect a drowning person is to detect the persons visible in the camera range, which is also done by background subtraction [45, 82, 104, 180]. In Figure 4 we give an example of a detection of the persons in a swimming pool using a background modelling algorithm. However, in this case aquatic environments have some added difficulties because it is not static [30].

4.3 Emergency Response/Assistance

After an emergency has already occurred, one of the most important tasks of emergency responders is to help injured people. However, depending on the magnitude of the disaster, it may be difficult to effectively localize all these persons so emergency responders may find it useful to use robots or UAVs to search for these persons. In order to automatically detect injured people, some of these robots use cameras together with computer vision algorithms. When considering disasters taking place on the ground one of the main characteristic of an injured person is that they are normally lying on the floor so these algorithms are trained to detect lying people. The main difference between them is the perspective from which these algorithms expect to get the input, depending if the camera is mounted on a terrestrial robot [22, 88, 141, 142] or on a UAV [7, 99, 131].

A fast and efficient response the in the moments after an emergency has occurred is critical to minimize damage. For that it is essential that emergency responders have situational awareness, which can be obtained through sensors in the damaged areas, by crowd sourcing applications, and from social media. However, extracting useful information from huge amounts of data can be a slow and tedious work. Therefore, organization of video data that facilitates the browsing on the data is very important [135]. Since during an emergency it is common to have data on the same event from different angles and perspectives, in [149] the authors combine interactive spatio-temporal exploration of videos and 3D reconstruction of scenes. In [72] the authors propose a way of representing the relationship among clips for easy video browsing. To extract relevant information from videos, in [38] the authors find recurrent regions in clips which is normally a sign of relevance. In [103] the authors introduce a framework of consecutive object detection from general to fine that automatically detects important features during an emergency. This also minimizes bandwidth usage, which is essential in many emergency situations where critical communications infrastructure collapses.

4.4 Emergency Understanding

Keeping record of emergencies is very important to analyse them and extract information in order to understand the factors that have contributed to provoke the emergency, how these factors could have been prevented, how the emergency could have been better assisted to reduce the damages or the effects of the emergency. The systems developed to with this objective should feed back into the rest of the systems in order to improve them.

| Location | Sensor | | | |
|----------------|---|--|---|---|
| | Monocular camera | Infrared/ Thermal camera | Stereo camera | Other |
| Fixed location | <ul style="list-style-type: none"> - \$50 to \$2000 - Up to 30 MP - Real time data acquisition - Limited area of visualization | <ul style="list-style-type: none"> - \$ 200 to \$ 6000 - Up to 1.2 MP - Up to 0.02°C sensitivity - Real time data acquisition - Limited area of visualization | <ul style="list-style-type: none"> - \$ 200 to \$ 2000 - Up to 30 MP - Real time data acquisition - Limited area of visualization | |
| Robot or UAV | <ul style="list-style-type: none"> - \$100 to \$2000 - Up to 30 MP - Portable - Bad quality data acquisition in real time - Limited autonomy | <ul style="list-style-type: none"> - \$250 to \$2000 - Up to 1.2 MP - Up to 0.02°C sensitivity - Portable - Bad quality data acquisition in real time | <ul style="list-style-type: none"> - \$ 250 to \$ 2000 - Up to 30 MP - Portable - Bad quality data acquisition in real time - Limited autonomy | |
| Satellite | | | | <ul style="list-style-type: none"> - Open data access through many research programs - Spatial resolution of around 10m - Can cover the entire globe - May have to wait a few days until the satellite covers the entire globe - Image acquisition is influenced by weather conditions like clouds |

Table 2: Summary of the advantages and disadvantages of the types of sensors and the location at which they can be placed.

5 Acquisition

Choosing an appropriate acquisition setup is one of the important design choices. To take this decision one must take into account the type of emergency addressed (see Section 3) and objectives to fulfill (see Section 4). One of the things to consider is the location where the sensor will be placed (e.g. in a fixed location, mounted on board a satellite, a robot, or a UAV, etc). Then, depending on the type of sensor we acquire different types of visual information: monocular, stereo, infrared, or thermal. Depending on the frequency in which the visual information is acquired we can obtain videos or still images. Finally, this section is summarized in Table 2 where we include the advantages and disadvantages of the types of sensors studied and the locations at which they can be placed. The type of sensors chosen for the system will have an important impact on the types of algorithms that can be applied, the type of information that can be extracted and the cost of the system, which is also many times a limitation.

5.1 Sensor location

The physical location where the sensor is installed determines the physical extent in the real world that will be captured by the sensor. The location of the sensor depends primarily on the expected physical extent and location of the emergency that is to be prevented, detected, assisted or understood.

In a fixed location: When the type of emergency considered has a limited extent and the location where it can occur (the area of risk) is predictable, a fixed sensor is normally installed. In many cases this corresponds to a monitoring camera mounted on a wall or some other environmental structure. These are the cheapest and most commonly used sensors. They are normally mounted in a high location, where no objects obstruct their view and pointing to the area of risk.

Static cameras have been used to detect drowning persons in swimming pools, as in this case the area of risk is limited to the swimming pool [26, 30, 42, 45, 82, 104, 180].

They have also been used to monitor dangerous areas on roads, inside vehicles either pointing to the driver to estimate drowsiness [31, 64, 94, 98], or pointing to the road to process the area in which it is being driven [54, 62, 67, 69, 107, 118]. Fixed cameras are also widely used in closed areas to detect falling persons [33, 79, 116, 130]. Finally, they can occasionally also be used to detect floods or fire in hotspots [16, 24, 96, 128].

On board robots or UAVs: For emergencies that do not take place in a specific area, or that need a more immediate attention, sensors can be mounted on mobile devices such as robots or UAVs. The main difference between these two types of devices is the perspective from which the data can be acquired: terrestrial in the case of a robot, and aerial for UAVs. On the one hand, the main advantage of these data sources is that they are flexible and support real time interaction with emergency responders. On the other hand, their main drawback is their limited autonomy due to battery capacity. Robots and UAVs have been used extensively for human rescue. In such cases, a visual sensor is mounted on board the autonomous machine which can automatically detect human bodies [7, 22, 99, 141, 142]. The authors of [66] use a UAV with several optical sensors to capture aerial images of the ground, estimate the soil moisture, and determine if an area suffers or is close to suffering drought.

On board satellites: For emergencies of a greater extent, such as drought, floods or fire, sometimes the use of sensors on board satellites is more convenient [66, 112, 113, 140]. By using satellites, it is possible to visualize the whole globe. However, the biggest drawback of this type of sensor is the frequency at which the data can be acquired since the satellites are in continuous movement and it may be necessary to wait up to a few days to obtain information about a specific area, as an example the Sentinel 2 which is a satellite in the Copernicus program for disaster management revisits every location within the same view angle every 5 days [43]. Due to this, the information acquired by this type of sensor is mainly used for assistance or understanding of an emergency rather than to prevent or detect it.

Multi-view cameras: In some scenarios, like in the street, it is useful to have views from different perspectives in order to increase the range or performance of the algorithm. However, using data from different cameras also entails some difficulties in combining the different views. Since the most common scenario where multiple cameras are present is in the street, this problem has been tackled in pedestrian detection [155] and traffic accidents [46].

5.2 Type of sensor

Another important decision to take when working with computer vision for emergency management is the type of opti-

cal sensor that should be used. When taking this decision it is important to take into account the technical characteristics of the available sensors, the type of data that they will capture, and the cost.

Monocular RGB or greyscale cameras: Monocular cameras are devices that transform light from the visible spectrum into an image. This type of visual sensor is the cheapest and therefore by far the most widely used in computer vision and monitoring systems. Although these sensors may not be the optimal in some scenarios (such as fire detection), due to their low cost and the experience of researchers on the type of images generated by such sensors, they are used in almost all emergencies [3, 42, 71, 96, 115, 151].

Stereo RGB or greyscale cameras: Stereo cameras are optical instruments that, similarly to monocular cameras, transform light from the visible spectrum into an image. The difference between stereo and monocular cameras is that stereo sensors have at least two lenses with separate image sensors. The combination of the output from the different lenses can be used to generate stereo images from which depth information can be extracted. This type of sensor are especially useful on board vehicles where the distance between the objects and the vehicle is essential to estimate if a collision is likely to occur and when [9]. They are also used on board of vehicles to detect lanes [37, 62].

Infrared/thermal cameras: Infrared or thermal cameras are devices that detect the infrared radiation emitted by all objects with a temperature above absolute zero [50]. Similarly to common cameras that transform visible light into images, thermal cameras convert infrared energy into an electronic signal which is then processed to generate a thermal image or video. This type of camera has the advantage that they eliminate the illumination problems that normal RGB or greyscale cameras have. They are of particular interest for emergency management scenarios such as fire detection since fire normally has a temperature higher than its environment.

It is also well known that humans have a body temperature that ranges between 36°C and 37°C in normal conditions. This fact can be exploited to detect humans, and has been used for pedestrian detection to prevent vehicle collisions with pedestrians [144, 161], to detect faces and pupils [94], and for fall detection for elderly monitoring [139]. Although not so extensively used, thermal and infrared cameras have also been used for military purposes, such as to detect firing guns [124] and to detect mines under ground [119, 138, 164].

One of the disadvantages of thermal cameras compared with other type of optical sensors is that their resolution is lower than many other visual sensors. For this reason some systems use thermal cameras in combination with other sensors. The authors of [66] use a thermal camera among other

visual sensors to study the surface soil moisture and the water balance of the ground, which is in turn used to estimate, predict and monitor drought.

In order to obtain depth information in addition to thermal information, it is also possible to use infrared stereo cameras such as in [14] where the authors use infrared cameras to detect pedestrians by detecting their corporal heat, and then estimate pedestrian position using stereo information.

5.3 Acquisition rate

Depending on the rate at which optical data is acquired we can either obtain video or still images. The main difference between video and a still image is that still images are taken individually while video is an ordered sequence of images taken at a specific and known rate. To be considered a video, the rate at which images are taken must be enough for the human eye to perceive movement as continuous.

Video: The main advantage of videos over images is that, thanks to the known acquisition rate, they provide temporal information which is invaluable for some tasks. Also, most of the systems studied in this survey are monitoring systems meant to automatically detect risky situations or emergencies. As these situations are rare but could happen at any moment and it is important to detect them as soon as possible, so it is necessary to have a high acquisition rate.

For the study of many of the emergencies considered in this survey it is necessary to have temporal information. For example, merely detecting a person lying on the ground in a single image is not sufficient to infer that that person has *fallen*. Many systems distinguish between a person lying on the floor and a person that has fallen by using information about the speed at which this person reached the final posture [102, 129]. Also, systems that estimate likelihood of a car crash must have temporal information to know at which speed the external objects are approaching the vehicle. To estimate driver drowsiness or fatigue, many systems take information into account such as the frequency at which the driver closes his eyes and for how long they remain closed [94].

Moreover, many algorithms developed for the detection of specific situations also require temporal information. In order to detect fire, most algorithms exploit the movement of the fire to differentiate it from the background. Similarly, for drowning person detection, many algorithms use temporal information to easily segment the person from the water [30, 42, 45, 104]. Also, many falling person detection systems use tracking algorithms to track the persons in the scene [68] or background subtraction algorithms to segment persons [116].

Still images: Still images, in contrast with video, do not give information about the movement of the objects in the scene and does not provide any temporal information. Although most of the algorithms that study the problem of fire detection use temporal information and therefore require video, David Van et al. [156] study a system that detects fire on still images to broaden the applicability of their system. In some cases due to the sensor used by the system it is not possible or convenient to obtain videos for example in systems that use satellite imagery [66, 112, 113, 140], also in systems that use UAVs where in many cases is more convenient to take still images instead of videos in order to effectively send the data in real time or get higher resolution imagery [22, 141].

5.4 Sensor cost

When implementing the final system, another important factor that should be taken into account before designing it is the cost from setting up the system. In general, the cost of the final system is going to be the cost of the visual sensor, the cost of the installation of the system and the cost of the data transmission. In this section we will do a brief study of the three sources of costs.

The most common visual sensors are monocular RGB and grayscale cameras. Since these sensors are so common and have a big demand, their technical features have improved greatly over time and their price has dramatically decreased. Therefore there exists a wide range of monocular cameras in the market with different technical features and prices. Depending on the quality of the device, its price can vary between approximately \$50 to \$2000. Although there also exists a wide variety of infrared cameras, these devices are less common and in general more expensive. Depending on the quality of the camera, its price fluctuates between \$200 and \$6000. To illustrate that, in the second row from Table 3 we give an example of a thermal camera of \$300 and its specifications. Finally, stereo cameras are the less common visual sensors and there is fewer variety of these devices. Their price oscillates between \$200 and \$2000. In the third row from Table 3 we present an example of an average stereo camera. However, it is fairly common to build a stereo device by attaching two monocular cameras which makes its price double the price of a monocular camera.

The cost of the installation of the system and the transmission of data varies greatly depending on the localization in which the sensor is installed. In this case, the cheapest installation would be at a fixed location, for which it would only be necessary to buy a holder to fix the sensor and acquire a broadband service in order to transmit the data in real time. On the other hand, it is also possible to install the sensor on a mobile device such as a robot or UAV. In this case it is necessary to first buy the device, whose price also




| Device | Description |
|---|--|
|  | Example of a commonly used UAV. This UAV has an integrated camera of 12 Mpx, a flying autonomy of 28 minutes, a maximum transmission distance of 5km a live view working frequency of 2.4GHz and a live view quality of 720P and 30fps. This device can be bought on the market for around \$1200. |
|  | Example of a thermal camera that can be attached to a smart phone. It has an VGA resolution a sensibility of 0.1°C and works with temperatures between -20°C to 120°C. A device if these characteristics can be bought for around \$300. |
|  | Example of a built in stereo camera of 10 MP resolution and a cost of \$600. |

Table 3: Some examples of devices used in computer vision for emergency management systems, their cost and specifications.

have big fluctuations depending on their quality and technical characteristics such as its autonomy, dimensions and stability. A low cost robot or UAV can be bought for around \$50 while high quality ones can cost around \$20000. Moreover, these devices require a complex transmission system since the sensor is going to be in continuous movement. In fact, since these systems are meant to acquire data over a limited period of time, it is common not to install any transmission data system but saving all the data in the device and processing it after the acquisition is over. It is also possible to install an antenna on the device that transmits these data in real time over 2.4GHz or 5.8GHz frequency. In the first row from Table 3 we show an example of a commonly used drone with an integrated camera and with a 2.4GHz antenna for live data transmission.

6 Algorithms

In this section we review the most popular algorithms that have been applied in computer vision applications for emergency situations. We have considered the nine different types of emergencies that we introduced in Section 3. When selecting the articles to review in this survey we have considered a total of 114 papers that propose a computer vision system for some emergency management. However, not all the types of emergencies have been equally studied in the literature nor in the survey, in Figure 5 we show a bar graph of the distribution of the articles among the different types of emergencies.

This section is divided in two main parts: feature extraction and machine learning algorithms. In the feature extrac-

tion part various approaches to extracting relevant information from the data are elaborated. Once that information is extracted, machine learning algorithms are applied to incorporate spatial consistency (e.g. Markov Random Fields), to model uncertainty (e.g. Fuzzy logic) and to classify the data to high-level semantic information, such as the presence of fire or an earthquake. We have focused on methods which are (or could be) applied to a wider group of applications in emergency management.

6.1 Feature extraction

Feature extraction in computer vision is the process of reducing the information in the data into a more compact representation which contains the relevant information for the task at hand. This is a common practice in computer vision algorithms related to emergency situations. We divide the features in three main groups: color, shape and temporal.

6.1.1 Color features

The understanding of emergency situations often involves the detection of materials and events which have no specific spatial extent or shape. Examples include water, fire, and smoke. However, these textures often have a characteristic color, and therefore color features can provide distinctive features.

Color analysis is a well established computer vision research field and has been broadly applied in applications in emergency management. It is frequently used to detect fire

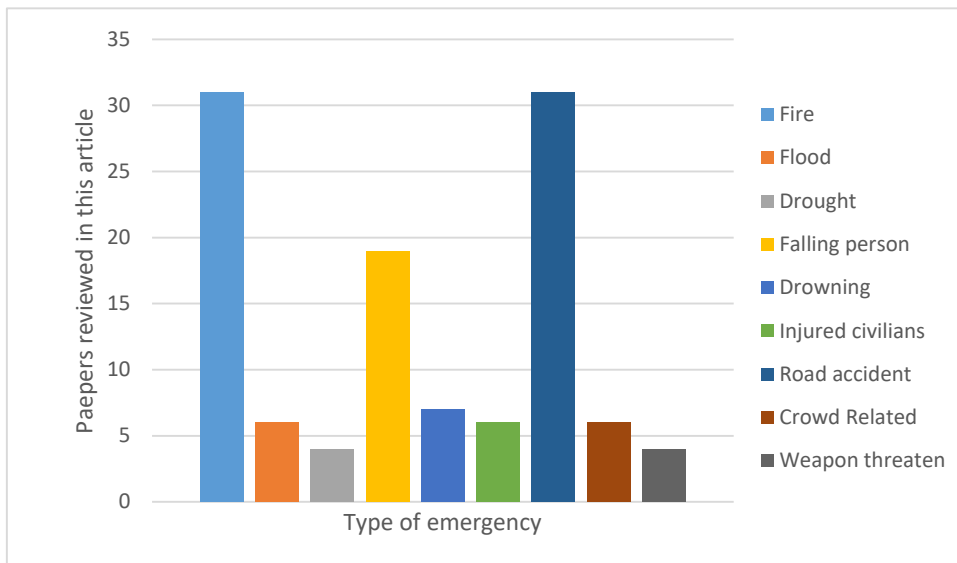


Fig. 5: Articles studied in this survey by type of emergency.

and smoke since they have characteristic colors and – although not so frequently – it has also been used to detect flooded regions. Fire color may vary depending on the combustion material and the temperature. However most flames are yellow, orange and especially red. Smoke tends to be dark or white and can be more or less opaque. Water in flooded regions, according to [16], presents a smooth dark surface or strong reflective spots much brighter than the surroundings.

One of the principle choices in color feature extraction is the color space or color model from which to compute the features. This choice influences characteristics of the color feature, including photometric invariance properties [55], device dependence, and whether it is perceptually uniform or not [74]. Here we will shortly review the most used color models and briefly describe some of their advantages and disadvantages.

The data produced by camera devices is typically in **RGB** color model. Due to the effortless availability of this color model and the separation of the red values, this color model has been often used for fire, smoke and flood detection [15, 27, 128, 151]. The main disadvantage of this color model is its lack of illumination invariance and dependence on the sensor. The **HSI** and **HSV** color models were proposed to resemble the color sensing properties of human vision and separate the luminance component from the chromatic ones. They have been applied to smoke detection [8], to detect light gray water [16] and to detect deep water [96]. Several systems combine RGB and HSI color models [29, 128, 174]. The **YCbCr** color model has three components, one of which encodes the luminance information and two encode the chromaticity. Having the luminance component sepa-

rated from chromatic components makes it robust to changes in illumination [24, 96]. Similarly to HSI and YCbCr, the **CIE Lab** color model makes it possible to separate the luminance information from the chromaticity. Moreover, it is device independent and perceptually uniform. This color model has been used to detect fire [23, 153] and to improve the background models for swimming pools [42].

6.1.2 Shape and texture features

There are elements which are very characteristic of some of the emergencies presented in this survey, such as smoke or flames to detect fire, pedestrians to prevent traffic collisions or persons to detect victims in disaster areas. Detecting these characteristic elements is common to many of the algorithms proposed for emergency management with computer vision. The primary goal of many of these algorithms is to prevent or detect human harm and therefore humans are a common characteristic element of these emergency scenarios. The histogram of oriented gradients (HOG) descriptor was introduced to detect pedestrians and is one of the most used shape descriptors to detect humans. Among the algorithms studied in this survey, HOG was used to detect persons in disaster areas [7, 141, 142], to detect pedestrians for on board vehicle cameras [144, 165] but also to characterize other objects such as flames, smoke and traffic signs [8, 160, 179] and as a descriptor for a tracking algorithm [51] in crowd event recognition. Local binary patterns is also a well known shape descriptor which is used to describe textures and was successfully applied to the task of smoke detection [148, 172].

6.1.3 Temporal features

Most of the input data used in emergency management comes in form of videos, which provide visual information as well as temporal information. Here we briefly summarize the main features which exploit the available temporal information in the data.

Wavelets: Flicker is an important characteristic of smoke and flames and has been used in several systems to detect fire and smoke in video data. Fourier analysis is the standard tool for analysis of spatial frequency patterns. The extension of frequency analysis to spatial-temporal signals was first studied by Haar [63] and later led to the wavelet transform [110].

In [151] the authors study the periodic behaviour of the boundaries of the moving fire-colored regions. The flickering flames can be detected by finding zero crossings of the wavelet transform coefficients which are an indicator of activity within the region. In [168] the flicker of smoke is detected computing the temporal high-frequency activity of foreground pixels by analysing the local wavelet energy. They also use frequency information to find smoke boundaries by detecting the decrease in high frequency content since the smoke is semi-transparent at its boundaries. In [21] the authors present an algorithm which uses a Discrete Wavelet Transform to analyse the blockwise energy of the image. The variance over time of this energy value was found to be an indicator of the presence or absence of smoke. In [57] the authors perform a vertical, horizontal and diagonal frequency analysis using the Stationary Wavelet Transform and its inverse. A characterization of smoke behaviour in the wavelet domain was done in [60] by decomposing the image into wavelets and analyzing its frequency information at various scales.

Optical flow: Optical flow estimates the motion of all the pixels of a frame with respect to the previous frame [10, 106]. The objective of algorithms based on optical flow is to determine the motion of objects in consecutive frames. Many techniques have been proposed in the literature to perform optical flow [18, 19, 52], however the most widely used technique is the one proposed by Lucas-Kanade [106] with some extensions [17].

Optical flow algorithms have been used in several computer vision systems for emergency management, such as for the detection of fire and smoke and anomalous situations on roads or in crowds. In the case of fire and smoke detection several approaches use the pyramidal Lucas-Kanade optical flow algorithm to determine the flow of candidate regions which likely to contain fire or smoke and to discard regions that do not present a high variation on the flow rate characteristic of the fire turbulence [128, 174]. The pyramidal Lucas-Kanade optical flow algorithm assumes that the objects in the video are rigid with a Lambertian surface and negligible changes in brightness but the drawback is that

these assumptions do not hold for fire and smoke. Addressing this problem, Kolesov et al. [91] proposed a modification to Lucas-Kanade in which they assume brightness and mass conservation. Optical flow has been also used to analyse the motion of crowds in order to determine abnormal and therefore potentially dangerous situations [4, 75, 115]. In [145] they use an optical flow technique to model the traffic dynamics of the road, however instead of doing a pixel by pixel matching they consider each individual car as a particle for the optical flow model.

Background modeling and subtraction: The separation of background from foreground is an important step in many of the applications for emergency management. Background modeling is an often applied technique for applications with static cameras (see Section 5.1) and static background. The aim of this method is to compute a representation of the video scene with no moving objects, a comprehensive review can be seen in [122].

In the context of emergency management, the method of Stauffer and Grimson [143] has been used to segment fire, smoke and persons from the background [5, 21, 104, 147, 153, 160]. Although not as popular, other background subtraction algorithms have been proposed in the literature [35, 77, 87, 100] and used in systems for emergency management [75, 89, 101, 137, 151]. An interesting application of background modeling was described in [83] in the context of abnormal event detection in emergency situations. They assume that vehicles involved in an accident or an abnormal situation remain still for a considerable period of time. Therefore an abnormal situation or accident can be detected in a surveillance video by comparing consecutive background models.

Tracking: Tracking in computer vision is the problem of locating moving objects of interest in consecutive frames of a video. The shape of tracked objects can be represented by points, primitive geometric shapes or the object silhouette [173]. Representing the object to be tracked as points is particularly useful when object boundaries are difficult to determine, for example when tracking crowds [75] where detecting and tracking each individual person may be infeasible. It is also possible to represent the tracked objects as primitive objective shapes, for example in [104] the authors use ellipses to represent swimmers for the detection of drowning incidents. It is also common to represent the tracked objects as rectangles, e.g. in swimmer detection [30] and traffic accident detection [83]. Finally, tracked objects can also be represented with the object's silhouette or contour [82].

Feature selection is crucial for tracking accuracy. The features chosen must be characteristic of the objects to be tracked and as unique as possible. In the context of fall detection, the authors of [68] use skin color to detect the head of the tracked person. Many tracking algorithms use motion as the main feature to detect the object of interest, for exam-

ple applications in traffic emergencies [71,83], fall detection [79,101], and drowning person detection [104]. Haar-like features [158] have been widely used in computer vision for object detection, and have been applied to swimmer detection [30]. Finally, descriptors such as Harris [65] or HOG [36] can also be useful for crowd analysis [51,75].

6.1.4 Convolutional features

Recently, convolutional features have led to a very good performance in analyzing visual imagery. Convolutional Neural Networks (CNN) were first proposed in 1990 by LeCun et al. [97]. However, due to the lack of training data and computational power, it was not until 2012 [93] that these networks started to be extensively used in the computer vision community.

CNNs have also been used in the recent years for some emergency challenges, mainly fire and smoke detection in images or video and anomalous event detection in videos. For the task of smoke detection, some researchers proposed novel networks to perform the task [48,167]. In [48] the authors propose a 9-layer convolutional network and they use the output of the last layer as a feature map of the regions containing fire, while in [167] the authors construct a CNN and use domain adaptation to alleviate the gap between their synthetic fire and smoke training data and their real test data.

Other researchers have concentrated on solutions built using existing networks [95,109]. In [109] the authors re-train some existing architectures, while in [95] the authors use already trained architectures to give labels to the images and then classify those labels into fire or not fire. For the task of anomaly detection, algorithms need to take temporal and spatial information into account. To handle this problem the authors in [32,114] propose to use a Long Short-Term Memory (LSTM) to predict the evolution of the video based on previous frames. If the reconstruction error between subsequent frames and the predictions is very high it is considered to be an anomaly. In [183] the authors consider only the regions of the video which have changed over the last period of time and send those regions through a spatial-temporal CNN. In [132] the authors pass a pixel-wise average of every pair of frames through a 4-layered fully connected network and an autoencoder trained on normal regions.

6.2 Machine learning

In the vast majority of the applications in emergency management machine learning algorithms are applied. They are used for classification and regression to high-level semantic information, to impose spatial coherence, as well as to estimate uncertainty of the models. Here we briefly discuss the most used algorithms.

6.2.1 Artificial Neural Networks

Artificial Neural Networks (ANN) have received considerable interest due to their wide range of applications and ability to handle problems involving imprecise and complex non-linear data. Moreover, their learning and prediction capabilities combine with impressive accuracy values for a wide range of problems. A neural network is composed of a large number of interconnected nonlinear computing elements or neurons organized in an input layer, a variable number of hidden layers and an output layer. The design of a neural network model requires three steps: selection of the statistical variables or features, selection of the number of layers (really the hidden layers) and nodes and selection of transfer function. So, the different statistical variables or features, the number of nodes and the selected transfer function, make the difference between the papers found showing applications of the ANN to prevention and detection of emergencies using computer vision and image processing techniques.

Applications of ANNs to fire and smoke detection and prevention can be seen in papers [91,169,174,175]. They mainly differ in the characteristics of the network. In [169] the authors extract features of the moving target which are normalized and fed to a single-layer artificial neural network with five inputs, one hidden layer with four processing elements, and one output layer with one processing element to recognize fire smoke. In [174] they propose a neural network to classify smoke features. In [91] they implement an ANN fire/smoke pixel classifier. The classifier is implemented as a single-hidden-layer neural network with twenty hidden units, with a softmax non-linearity in the hidden layer. The method proposed in [175] describes a neural network classifier with a hidden layer is trained and used for discrimination of smoke and non-smoke objects. The output layer uses a sigmoid transfer function.

In the system presented in [2] a neural network of perceptrons was used to classify falls against the set of predefined situations. Here, the neural network analyses the binary map image of the person and identifies which plausible situation the person is at any particular instant in time.

In general, the usage of ANNs provide a performance improvement compared to other learning methods [34]. The results rely highly on the selected statistical values for training. If the selection of statistical values is not appropriate, the results on the accuracy might be lower and simpler machine learning algorithm may be preferred.

Deep Learning ANN algorithms that consist of more than one hidden layer are usually called *deep* learning algorithms. Adding more layers to the network increases the complexity of the algorithm making it more suitable to learn more complex tasks. However, increasing the layers of the network comes at a higher computational cost and a higher probability to overfit or converge to a local minima. Nevertheless, the

advances in computers and the increase in training data has facilitated the research towards deeper models which have proven to improve the results.

In [34] the authors propose a multi-layered ANN for smoke features classification. In the works [2, 47, 80, 139] we find applications of ANNs to prevent and detect single-person emergencies. A neural network is used to classify falls in [139] using vertical velocity estimates derived either directly from the sensor data or from the tracker. The network described in [81], with four hidden layers, is used in the classifier design to do posture classification that can be used to determine if a person has fallen. In [47] the authors monitor human activities using a multilayer perceptron with focus on detecting three types of fall.

6.2.2 Support Vector Machines

Support vector machines (SVMs) are a set of supervised machine learning algorithms based on statistical learning and risk minimization which are related to classification and regression problems. Given a set of labeled samples a trained SVM creates an hyperplane or a set of them, in a high dimensional space, that optimally separates the samples between classes maximizing the separation margin among classes. Many computer vision systems related to emergency situations can be considered as a binary classification problems: fire or no fire, fall or no fall, crash or no crash. To solve this classification problem many researches make use of SVMs [116, 144, 148, 153] based on the classification performance that they can reach, their effectiveness to work on high dimensional spaces and their ability to achieve good performance with few data.

In more complex systems where it is difficult to predict the type of emergencies that may occur, for example in systems that supervise crowded areas or traffic, algorithms that consider abnormal behaviours as potential emergencies have been proposed [123, 163]. A difficulty of these kind of systems is that they can only be trained with samples of one class (normal scenes). For these kind of problems one-class SVMs, which find a region of a high dimensional space that contains the majority of the samples, are well suited.

6.2.3 Hidden Markov Models

Hidden Markov Models (HMMs), which are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture and action recognition [170], are a type of stochastic state transition model [125]. HMMs make it possible to deal with time-sequence data and can provide time-scale invariability in recognition. Therefore it is one of the preferred algorithms to analyze emergency situations subject to time variability.

To detect smoke regions from video clips, a dynamic texture descriptor is proposed in [172]. The authors propose a texture descriptor with Surfacelet transformation ([105]) and Hidden Markov Tree model (HMT). Examples of use of Markov models to study human-caused emergencies that affect a single person can be seen in the following works [26, 42, 79, 147, 150], covering fall detection and detection behavior of a swimmer and drownings. Hidden Markov Models have been applied on monitoring emergency situations in crowds by learning patterns of normal crowd behaviour in order to identify unusual or emergency events, as can be seen in [4, 6]. In these two works Andrade and his team presents similar algorithms in order to detect emergency events in crowded scenarios, and based on optical flow to extract information about the crowd behaviour and the Hidden Markov Models to detect abnormal events in the crowd.

Töreyn and his team in [150] use Hidden Markov Models to recognize possible human motions in video. Also they use audio track with a three-state HMM to distinguish a person sitting on a floor from a person stumbling and falling. In [147] the motion analysis is performed by means of a layered hidden Markov model (LHMM). A fall detection system based on motion and posture analysis with surveillance video is presented in [79]. The authors propose a context-based fall detection system by analyzing human motion and posture using a discrete hidden Markov model. The problem of recognition of several specific behaviours of a swimmer (including backstroke, breaststroke, freestyle and butterfly, and abnormal behaviour) in a swimming pool using Hidden Markov Models is addressed in [26]. The algorithm presented in [42] detects water crises within highly dynamic aquatic environments where partial occlusions are handled using a Markov Random Field framework.

Hidden Markov Models have been applied on monitoring emergency situations in crowds by learning patterns of normal crowd behavior in order to identify unusual or emergency events, as can be seen in [4, 6]. In these two works Andrade and his team presents similar algorithms in order to detect emergency events in crowded scenarios, and based on optical flow to extract information about the crowd behavior and the Hidden Markov Models to detect abnormal events in the crowd.

6.2.4 Fuzzy logic

Fuzzy sets theory and fuzzy logic are powerful frameworks for performing automated reasoning, and provide a principled approach to address the inherent uncertainty related to modeling; an aspect which relevant in many emergency management application. Fuzzy set theory, originally introduced by Lofti A. Zadeh in 1965 [177], is an extension of classical set theory where to each element x from a set A is associated with a value between 0 and 1, representing its

degree of belonging to the set A . One important branch of the fuzzy set theory is fuzzy logic [178]. An inference engine operators on rules that are structured in an IF-THEN rules, are presented. Rules are constructed from linguistic variables. A fuzzy approach has the advantage that the rules and linguistic variables are understandable and simplify addition, removal, and modification of the system's knowledge.

In the papers [3] and [147] the authors proposed a multi-camera video-based methods for monitoring human activities, with a particular interest to the problem of fall detection. They use methods and techniques from fuzzy logic and fuzzy sets theory. Thome et al. [147] apply a fuzzy logic fusion process that merges features of each camera to provide a multi-view pose classifier efficient in unspecified conditions (viewpoints). In [3], they construct a three-dimensional representation of the human from silhouettes acquired from multiple cameras monitoring the same scene, called voxel person. Fuzzy logic is used to determine the membership degree of the person to a pre-determined number of states at each image. A method, based on temporal fuzzy inference curves representing the states of the voxel person, is presented for generating a significantly smaller number of rich linguistic summaries of the humans state over time with the aim to detect of fall detection.

Another problem addressed with methods and techniques from theory of fuzzy sets is fire prevention by smoke detection. It has been observed that the bottom part of a smoke region is less mobile than top part over time; this is called swaying motion of smoke. In [162] a method to detect early smoke in video is presented, using swaying motion and a diffusion feature. In order to realized data classification based on information fusion, Choquet fuzzy integral (see [59]) was adopted to extract dynamic regions from video frames, and then, a swaying identification algorithm based on centroid calculation was used to distinguish candidate smoke region from other dynamic regions. Finally, Truong et al. [153] presented a four-stage smoke detection algorithm, where they use a fuzzy c-means algorithm method to cluster candidate smoke regions from the moving regions.

6.3 Algorithmic trends in emergency management

To analyze the usage of the discussed algorithms across emergency types, we provide a more in depth analysis of the three most studied emergencies: fire, road accidents, and human fall. In Figure 6(a-c) we break down the algorithms that have been used for these three types of emergencies. As can be seen in the figures, for the three types of emergencies, the most commonly applied algorithms are the ones that use temporal features. For fire and human fall management background subtraction is the most popular algorithm since most times the most discriminative feature of the objects of

interest in this scenarios move on top of a static background. In the case of road accident management it is common to use non-static cameras (such as cameras on board a vehicle) and therefore the background is not static and background subtraction algorithms are not useful. However, temporal features are used to perform optical flow or tracking in order to determine the trajectory of the objects in the road or close by. For fire color features are very distinctive and therefore more researched.

7 Discussion

There are many video monitoring systems whose aim is to monitor various types of emergency scenarios. Since continuous human vigilance of these recordings is infeasible, most of this data is used *a posteriori*. Although this information is very useful for the later understanding of the event or as an incriminatory proof, many researchers study the possibility of exploiting these huge amounts of data and analysing them in real time with the hope of preventing some of these emergencies or to facilitate a faster or more efficient response. In this survey we reviewed and analyzed the progress of computer vision covering many types of emergencies with the goal of unveiling the overlap across emergency research in computer vision, especially in the hardware and algorithms developed, making it easier to researchers to find relevant works.

In order to restrict the scope of this survey we first gave a description of what we consider emergency situations. Taking into account the description provided, we restrict ourselves to emergency scenarios that have been studied in the context of computer vision. Taking all these considerations into account we identified eleven different emergency scenarios, studied in computer vision, which we classified into natural emergencies and emergencies caused by humans. After a thorough analysis of the state-of-the-art algorithms on these topics, we inferred that there are four main axes that have to be taken into account when creating computer vision systems to help in emergency scenarios: the type of emergency that is going to be addressed, the monitoring objective, the hardware and sensors needed, and the algorithms used.

When talking about the monitoring objective, we have seen that prevention through computer vision techniques can only be done when there is visual evidence of risk factors such as for fire detection using smoke features, car accidents by detecting pedestrians, driver drowsiness and unusual behaviours, and finally fall prevention falls by detecting slips. The most common objective is emergency detection, which has been studied in all the emergencies studied in this survey. It is also common to use computer vision to detect humans in disaster areas to assist emergency responders. Fi-

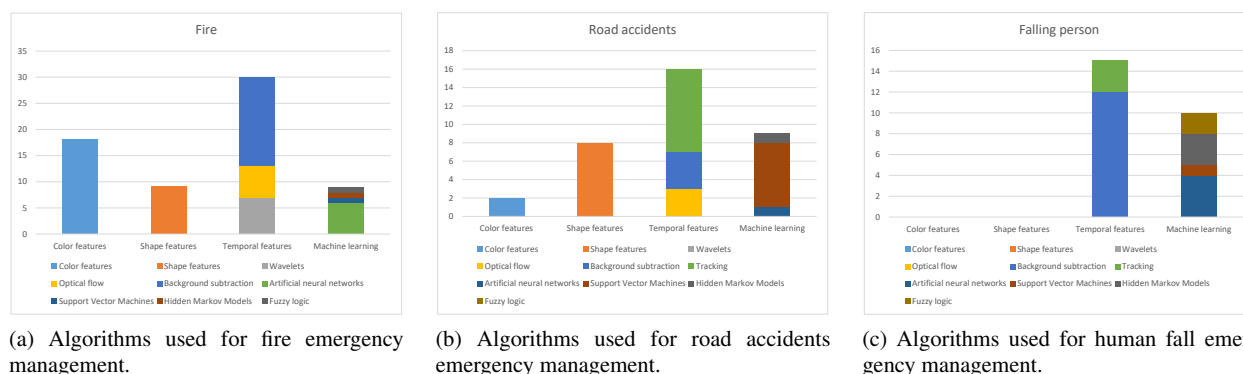


Fig. 6: Overview of the algorithms that have been used to manage the three most studied emergency scenarios.

nally, computer vision has not played a role from the emergency understanding point of view.

From all the acquisition methods studied in Section 5 we have concluded that the most common sensors are monocular RGB or grayscale cameras – mainly for their low cost – and that they are normally situated in fixed locations. However, we see a trend in the use of UAVs which are becoming cheaper and provide great flexibility through their mobility.

Finally, as discussed in Section 6.3, it is very difficult to draw conclusions on algorithmic trends for every emergency since some of them have little impact in computer vision. However, there is a clear tendency towards using color and temporal features for fire emergencies, a clear preference for temporal features in road accidents, and tracking algorithms and background subtraction techniques are the most common when detecting falling person events.

This up-to-date survey will be important for helping readers obtain a global view. At the same time they can find novel ideas of how computer vision can be helpful to prevent, respond and recover from emergency scenarios, and learn about some required hardware and relevant algorithms that are developed to solve these problems.

It is extremely difficult to compare different algorithms proposed in the literature for solving similar problems due to the lack of unified benchmarks. From all the papers studied during this review, over 60% used non-publicly available datasets. In order to facilitate a common benchmark to compare algorithms, in Table 4 we report some of the most important publicly available datasets used among the emergencies studied in this survey. Note that some pedestrian datasets are shared among emergencies since detecting humans is considered useful in more than one emergency.

In general, we have seen a tendency to create more complex systems that combine different algorithms or algorithms of a higher computational complexity which in many cases could not have been feasible with the previous generations of computers. Among these computationally complex al-

gorithms, the emerging techniques that stand out are those based on Convolutional Neural Networks (CNN) and Deep Learning [58, 84, 181]. The field of computer vision is experiencing rapid change since the rediscovery of CNNs. These algorithms have already been introduced in some emergency scenarios studied in this survey and we believe that the impact of CNNs will grow over the years in the field of emergency management with computer vision. Moreover, with the decreasing cost of visual acquisition devices such as infrared cameras or stereo cameras, there is a trend of combining different information sources which can help to increase accuracy rates and decrease false alarms. We also believe that, with the help of this survey and further study of the overlap that exists between the algorithms developed to detect and study different types of emergency situations, it will be possible in the future to create system that can cover more than one type of emergency.

References

1. Abdallah, S., Burnham, G.: Public health guide for emergencies. Boston: The Johns Hopkins School of Hygiene and Public Health and The International Federation of Red Cross and Red Crescent Societies (2000)
2. Alhimal, L., Zedan, H., Al-Bayatti, A.: The implementation of an intelligent and video-based fall detection system using a neural network. *Applied Soft Computing* **18**, 59 – 69 (2014)
3. Anderson, D., Luke, R.H., Keller, J.M., Skubic, M., Rantz, M., Aud, M.: Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Computer vision and image understanding* **113**(1), 80–89 (2009)
4. Andrade, E.L., Blunsden, S., Fisher, R.B.: Hidden Markov models for optical flow analysis in crowds. In: 18th International Conference on Pattern Recognition (ICPR'06), vol. 1, pp. 460–463. IEEE (2006)
5. Andrade, E.L., Blunsden, S., Fisher, R.B.: Modelling crowd scenes for event detection. In: 18th International Conference on Pattern Recognition (ICPR'06), vol. 1, pp. 175–178. IEEE (2006)
6. Andrade, E.L., Fisher, R.B., Blunsden, S.: Detection of emergency events in crowded scenes. In: *Crime and Security*, 2006.

| Type of emergency | Public datasets | General comments | Url |
|-------------------|--|--|---|
| Fire | VISOR smoke dataset | Videos with evidence of smoke. | http://imagelab.ing.unimore.it/visor |
| | Fire dataset from the Bilkent University | Videos with evidence of smoke and fire. | http://signal.ee.bilkent.edu.tr/VisiFire/ |
| | MESH database of news content | Instances of catastrophe related videos from the Deutsche Welle broadcaster with several news related to fires in a variety of conditions. | http://www.mesh-ip.eu/ |
| | FASTData Fire and smoke dataset | Collection of fire images and videos. Clips containing evidence of fire and smoke in several scenarios. | http://fire.nist.gov/fastdata http://signal.ee.bilkent.edu.tr/VisiFire/Demo |
| Fall detection | CIRL fall detection dataset | Indoor videos for action recognition and fall detection, not available for commercial purposes. | http://www.derektanderson.com/fallrecognition/datasets.html |
| | MMU fall detection dataset | 20 indoor videos including 38 normal activities and 29 different falls | http://foe.mmu.edu.my/digitalhome/FallVideo.zip |
| | Multiple camera fall dataset | 24 videos of falls and fall confounding situations recorded with 8 cameras in different angles. | http://www.iro.umontreal.ca/~labimage/Dataset/ |
| Injured civilians | NICTA dataset | Contains a total of 25551 unique pedestrians. | https://research.csiro.au/data61/automap-datasets-and-code |
| | MIT pedestrian dataset | 924 images of 64x128 containing pedestrians. | http://cbcl.mit.edu/software-datasets/PedestrianData.html |
| Road accident | MIT pedestrian dataset | 924 images of 64x128 containing pedestrians. | http://cbcl.mit.edu/software-datasets/PedestrianData.html |
| | INRIA pedestrian dataset | Images of pedestrians and negative images. | http://pascal.inrialpes.fr/data/human/ |
| | CAVIAR dataset | Sequences containing pedestrians, also suitable for action recognition. | http://groups.inf.ed.ac.uk/vision/CAVIAR |
| | Trajectory based anomalous event detection | Synthetic and real-world data. | http://ieeexplore.ieee.org/document/4633642/ |
| | The German Traffic Sign detection benchmark | 900 images of roads with traffic sign ground truth. | http://benchmark.ini.rub.de |
| | Thermal pedestrian database | 10 sequences containing pedestrians recorded with a thermal camera. | http://vcipl-okstate.org/pbvs/bench/ |
| | TUD Brussels and TUD Paris, Multi-Cue Onboard Pedestrian Detection | Images of pedestrians taken by on-board cameras. | https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/people-detection-pose-estimation-and-tracking |
| Crowd related | Simulation of Crowd Problems for Computer Vision | Approach for generating video evidence of dangerous situations in crowded scenes. | https://www.icsa.inf.ed.ac.uk/publications/online/0493.pdf |
| | PETS dataset | Sequences containing different crowd activities. | http://www.cvg.reading.ac.uk/PETS2009/a.html |
| | Unusual crowd activity dataset | Clip with unusual crowd activities. | http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi |
| | USCD anomaly detection dataset | Clips of the street with stationary camera. | http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm |
| | UMN monitoring human activity dataset | Videos of crowded scenes. | http://mha.cs.umn.edu/proj_recognition.shtml#crowd_count |

Table 4: Recompilation of the publicly available datasets used in the papers reviewed in this manuscript. Only about 40% of the papers reviewed used datasets publicly available.

The Institution of Engineering and Technology Conference on, pp. 528–533. IET (2006)

7. Andriluka, M., Schnitzspan, P., Meyer, J., Kohlbrecher, S., Petersen, K., Von Stryk, O., Roth, S., Schiele, B.: Vision based victim detection from unmanned aerial vehicles. In: Intelligent

- Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, pp. 1740–1747. IEEE (2010)
8. Barmoutis, P., Dimitropoulos, K., Grammalidis, N.: Smoke detection using spatio-temporal analysis, motion modeling and dynamic texture recognition. In: 2014 22nd European Signal Processing Conference (EUSIPCO), pp. 1078–1082. IEEE (2014)
 9. Barth, A., Franke, U.: Tracking oncoming and turning vehicles at intersections. In: Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on, pp. 861–868. IEEE (2010)
 10. Beauchemin, S.S., Barron, J.L.: The computation of optical flow. *ACM computing surveys (CSUR)* **27**(3), 433–466 (1995)
 11. Bejjiga, M.B., Zeggada, A., Nouffidj, A., Melgani, F.: A convolutional neural network approach for assisting avalanche search and rescue operations with UAV imagery. *Remote Sensing* **9**(2), 100 (2017)
 12. Below, R., Wirtz, A., Guha-Sapir, D.: Disaster category classification and peril terminology for operational purposes. Brussels: CRED/Munich: MunichRe Foundation (2009)
 13. Benenson, R., Omran, M., Hosang, J., Schiele, B.: Ten years of pedestrian detection, what have we learned? In: Computer Vision-ECCV 2014 Workshops, pp. 613–627. Springer (2014)
 14. Bertozzi, M., Broggi, A., Lasagni, A., Rose, M.: Infrared stereo vision-based pedestrian detection. In: IEEE Proceedings. Intelligent Vehicles Symposium, 2005., pp. 24–29. IEEE (2005)
 15. Borges, P.V.K., Izquierdo, E.: A probabilistic approach for vision-based fire detection in videos. *IEEE transactions on circuits and systems for video technology* **20**(5), 721–731 (2010)
 16. Borges, P.V.K., Mayer, J., Izquierdo, E.: A probabilistic model for flood detection in video sequences. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pp. 13–16. IEEE (2008)
 17. Bouguet, J.Y.: Pyramidal implementation of the affine Lucas-Kanade feature tracker description of the algorithm. *Intel Corporation* **5**(1-10), 4 (2001)
 18. Brox, T., Bruhn, A., Papenberger, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: European conference on computer vision, pp. 25–36. Springer (2004)
 19. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE transactions on pattern analysis and machine intelligence* **33**(3), 500–513 (2011)
 20. Buch, N., Velastin, S.A., Orwell, J.: A review of computer vision techniques for the analysis of urban traffic. *Intelligent Transportation Systems, IEEE Transactions on* **12**(3), 920–939 (2011)
 21. Calderara, S., Piccinini, P., Cucchiara, R.: Smoke detection in video surveillance: a MoG model in the wavelet domain. In: International conference on computer vision systems, pp. 119–128. Springer (2008)
 22. Castillo, C., Chang, C.: A method to detect victims in search and rescue operations using template matching. In: Safety, Security and Rescue Robotics, Workshop, 2005 IEEE International, pp. 201–206. IEEE (2005)
 23. Celik, T.: Fast and efficient method for fire detection using image processing. *ETRI journal* **32**(6), 881–890 (2010)
 24. Celik, T., Demirel, H.: Fire detection in video sequences using a generic color model. *Fire Safety Journal* **44**(2), 147–158 (2009)
 25. Çetin, A.E., Dimitropoulos, K., Gouverneur, B., Grammalidis, N., Günay, O., Habiboglu, Y.H., Töreyn, B.U., Verstockt, S.: Video fire detection—review. *Digital Signal Processing* **23**(6), 1827–1843 (2013)
 26. Chen, H.L., Tsai, M., Chan, C.: A Hidden Markov model-based approach for recognizing swimmer’s behaviors in swimming pool. In: 2010 International Conference on Machine Learning and Cybernetics, vol. 5, pp. 2459–2465. IEEE (2010)
 27. Chen, J., He, Y., Wang, J.: Multi-feature fusion based fast video flame detection. *Building and Environment* **45**(5), 1113–1122 (2010)
 28. Chen, J., Kwong, K., Chang, D., Luk, J., Bajcsy, R.: Wearable sensors for reliable fall detection. In: Engineering in Medicine and Biology Society, pp. 3551–3554 (2005)
 29. Chen, T.H., Wu, P.H., Chiou, Y.C.: An early fire-detection method based on image processing. In: Image Processing, 2004. ICIP’04. 2004 International Conference on, vol. 3, pp. 1707–1710. IEEE (2004)
 30. Chen, W.H., Cho, P.C., Fan, P.L., Yang, Y.W.: A framework for vision-based swimmer tracking. In: Uncertainty Reasoning and Knowledge Engineering (URKE), 2011 International Conference on, vol. 1, pp. 44–47. IEEE (2011)
 31. Choi, I.H., Kim, Y.G.: Head pose and gaze direction tracking for detecting a drowsy driver. In: Big Data and Smart Computing (BIGCOMP), 2014 International Conference on, pp. 241–244. IEEE (2014)
 32. Chong, Y.S., Tay, Y.H.: Abnormal event detection in videos using spatiotemporal autoencoder. In: International Symposium on Neural Networks, pp. 189–196. Springer (2017)
 33. Chua, J.L., Chang, Y.C., Lim, W.K.: A simple vision-based fall detection technique for indoor video surveillance. *Signal, Image and Video Processing* **9**(3), 623–633 (2015)
 34. Chunyu, Y., Jun, F., Jinjun, W., Yongming, Z.: Video fire smoke detection using motion and color features. *Fire Technology* **46**(3), 651–663 (2010)
 35. Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P., et al.: A system for video surveillance and monitoring. Tech. rep., Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University (2000)
 36. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 1, pp. 886–893. IEEE (2005)
 37. Danescu, R., Nedevschi, S.: Probabilistic lane tracking in difficult road scenarios using stereovision. *IEEE Transactions on Intelligent Transportation Systems* **10**(2), 272–282 (2009)
 38. Diem, L., Zaharieva, M.: Video content representation using recurring regions detection. In: International Conference on Multimedia Modeling, pp. 16–28. Springer (2016)
 39. Martínez-de Dios, J.R., Arrue, B.C., Ollero, A., Merino, L., Gómez-Rodríguez, F.: Computer vision techniques for forest fire perception. *Image and vision computing* **26**(4), 550–562 (2008)
 40. Doerr, K.U., Hutchinson, T.C., Kuester, F.: A methodology for image-based tracking of seismic-induced motions. In: Proc. SPIE, vol. 5758, pp. 321–332 (2005)
 41. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: An evaluation of the state of the art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **34**(4), 743–761 (2012)
 42. Eng, H., Toh, K., Kam, A.H., Wang, J., Yau, W.Y.: An automatic drowning detection surveillance system for challenging outdoor pool environments. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, pp. 532–539. IEEE (2003)
 43. ESA: Sentinel 2 mission details (2014). URL <https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/sentinel-2>
 44. FATE: New fall detector arrives this september 2015 (2015). URL <http://fate.upc.edu/index.php>
 45. Fei, L., Xueli, W., Dongsheng, C.: Drowning detection based on background subtraction. In: Embedded Software and Systems, 2009. ICES’09. International Conference on, pp. 341–343. IEEE (2009)
 46. Forczmański, P., Nowosielski, A.: Multi-view data aggregation for behaviour analysis in video surveillance systems. In: International Conference on Computer Vision and Graphics, pp. 462–473. Springer (2016)

47. Foroughi, H., Aski, B.S., Pourreza, H.: Intelligent video surveillance for monitoring fall detection of elderly in home environments. In: *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on*, pp. 219–224 (2008)
48. Frizzi, S., Kaabi, R., Bouchouicha, M., Ginoux, J.M., Moreau, E., Fnaiech, F.: Convolutional neural network for video fire and smoke detection. In: *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*, pp. 877–882. IEEE (2016)
49. Fu, Q., Banerjee, A., Liess, S., Snyder, P.K.: Drought detection of the last century: An MRF-based approach. In: *SDM*, pp. 24–34. SIAM (2012)
50. Gade, R., Moeslund, T.B.: Thermal cameras and applications: a survey. *Machine vision and applications* **25**(1), 245–262 (2014)
51. Garate, C., Bilinsky, P., Bremond, F.: Crowd event recognition using HOG tracker. In: *Performance Evaluation of Tracking and Surveillance (PETS-Winter), 2009 Twelfth IEEE International Workshop on*, pp. 1–6. IEEE (2009)
52. Gautama, T., Van Hulle, M.: A phase-based approach to the estimation of the optical flow field using spatial filtering. *IEEE Transactions on Neural Networks* **13**(5), 1127–1136 (2002)
53. Gavrilu, D.M., Munder, S.: Multi-cue pedestrian detection and tracking from a moving vehicle. *International journal of computer vision* **73**(1), 41–59 (2007)
54. Geronimo, D., Lopez, A.M., Sappa, A.D., Graf, T.: Survey of pedestrian detection for advanced driver assistance systems. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **37**(7), 1239–1258 (2009)
55. Gevers, T., Gijzenij, A., van de Weijer, J., Geusebroek, J.: *Color in Computer Vision: Fundamentals and Applications*. Wiley (2012)
56. Gomez-Rodriguez, F., Arrue, B., Ollero, A.: Smoke monitoring and measurement using image processing: application to forest fires. In: *AeroSense 2003*, pp. 404–411. International Society for Optics and Photonics (2003)
57. Gonzalez-Gonzalez, R., Alarcon-Aquino, V., Romero, R.R., Starostenko, O., Rodriguez-Asomoza, J., Ramirez-Cortes, J.: Wavelet-based smoke detection in outdoor video sequences. In: *Circuits and Systems (MWSCAS), 2010 53rd IEEE International Midwest Symposium on*, pp. 383–387. IEEE (2010)
58. Gope, S., Sarkar, S., Mitra, P., Ghosh, S.: *Early Prediction of Extreme Rainfall Events: A Deep Learning Approach*, pp. 154–167. Springer International Publishing, Cham (2016)
59. Grabisch, M., Murofushi, T., Sugeno, M.: *Fuzzy Measures and Integrals: Theory and Applications*. Studies in Fuzziness and Soft Computing. Physica-Verlag HD (2000)
60. Gubbi, J., Marusic, S., Palaniswami, M.: Smoke detection in video using wavelets and support vector machines. *Fire Safety Journal* **44**(8), 1110–1115 (2009)
61. Günay, O., Taşdemir, K., Töreyn, B.U., Çetin, A.E.: Fire detection in video using LMS based active learning. *Fire technology* **46**(3), 551–577 (2010)
62. Guo, C., Mita, S., McAllester, D.: Robust road detection and tracking in challenging scenarios based on Markov random fields with unsupervised learning. *Intelligent Transportation Systems, IEEE Transactions on* **13**(3), 1338–1354 (2012)
63. Haar, A.: Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen* **69**(3), 331–371 (1910)
64. Hachisuka, S., Ishida, K., Enya, T., Kamijo, M.: Facial expression measurement for detecting driver drowsiness. In: *Engineering Psychology and Cognitive Ergonomics*, pp. 135–144. Springer (2011)
65. Harris, C., Stephens, M.: A combined corner and edge detector. In: *Proceedings of the 4th Alvey Vision Conference*, pp. 147–151 (1988)
66. Hassan-Esfahani, L., Torres-Rua, A., Jensen, A., McKee, M.: Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. *Remote Sensing* **7**(3), 2627–2646 (2015)
67. Hayashi, T., Yamada, K.: Predicting unusual right-turn driving behavior at intersection. In: *2009 IEEE Intelligent Vehicles Symposium* (2009)
68. Hazelhoff, L., Han, J., et al.: Video-based fall detection in the home using principal component analysis. In: *Advanced Concepts for Intelligent Vision Systems*, pp. 298–309. Springer (2008)
69. Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., Igel, C.: Detection of traffic signs in real-world images: The german traffic sign detection benchmark. In: *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pp. 1–8. IEEE (2013)
70. Hu, S., Zheng, G.: Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Systems with Applications* **36**(4), 7651–7658 (2009)
71. Hu, W., Xiao, X., Fu, Z., Xie, D., Tan, T., Maybank, S.: A system for learning statistical motion patterns. *IEEE transactions on pattern analysis and machine intelligence* **28**(9), 1450–1464 (2006)
72. Huang, H., Liu, H., Zhang, L.: Videoweb: space-time aware presentation of a videoclip collection. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* **4**(1), 142–152 (2014)
73. Hutchinson, T.C., Kuester, F.: Monitoring global earthquake-induced demands using vision-based sensors. *IEEE Transactions on Instrumentation and Measurement* **53**(1), 31–36 (2004)
74. Ibraheem, N.A., Hasan, M.M., Khan, R.Z., Mishra, P.K.: Understanding color models: a review. *ARNP Journal of Science and Technology* **2**(3), 265–275 (2012)
75. Ihaddadene, N., Djeraba, C.: Real-time crowd motion analysis. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4. IEEE (2008)
76. IHSTechnology: 245 million video surveillance cameras installed globally in 2014 (2015). URL <https://technology.ihs.com/532501/245-million-video-surveillance-cameras-installed-globally-in-2014>
77. Izadi, M., Saeedi, P.: Robust region-based background subtraction and shadow removing using color and gradient information. In: *2008 19th International Conference on Pattern Recognition*, pp. 1–5 (2008)
78. Jiang, F., Yuan, J., Tsaftaris, S.A., Katsaggelos, A.K.: Anomalous video event detection using spatiotemporal context. *Computer Vision and Image Understanding* **115**(3), 323–333 (2011)
79. Jiang, M., Chen, Y., Zhao, Y., Cai, A.: A real-time fall detection system based on HMM and RVM. In: *Visual Communications and Image Processing (VCIP), 2013*, pp. 1–6. IEEE (2013)
80. Juang, C.F., Chang, C.M.: Human body posture classification by a neural fuzzy network and home care system application. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* **37**(6), 984–994 (2007)
81. Juang, C.F., Lin, C.T.: An online self-constructing neural fuzzy inference network and its applications. *IEEE Transactions on Fuzzy Systems* **6**(1), 12–32 (1998)
82. Kam, A.H., Lu, W., Yau, W.Y.: A video-based drowning detection system. In: *European Conference on Computer Vision*, pp. 297–311. Springer (2002)
83. Kamijo, S., Matsushita, Y., Ikeuchi, K., Sakauchi, M.: Traffic monitoring and accident detection at intersections. *Intelligent Transportation Systems, IEEE Transactions on* **1**(2), 108–118 (2000)
84. Kang, B., Choo, H.: A deep-learning-based emergency alert system. *ICT Express* **2**(2), 67–70 (2016)
85. Kanistras, K., Martins, G., Rutherford, M.J., Valavanis, K.P.: Survey of unmanned aerial vehicles (UAVs) for traffic monitoring.

- In: Handbook of Unmanned Aerial Vehicles, pp. 2643–2666. Springer (2015)
86. Ke, S.R., Thuc, H.L.U., Lee, Y.J., Hwang, J.N., Yoo, J.H., Choi, K.H.: A review on video-based human activity recognition. *Computers* **2**(2), 88–131 (2013)
 87. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Real-time foreground–background segmentation using codebook model. *Real-time imaging* **11**(3), 172–185 (2005)
 88. Kleiner, A., Kummerle, R.: Genetic MRF model optimization for real-time victim detection in search and rescue. In: Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on, pp. 3025–3030. IEEE (2007)
 89. Ko, B., Cheong, K.H., Nam, J.Y.: Early fire detection algorithm based on irregular patterns of flames and hierarchical bayesian networks. *Fire Safety Journal* **45**(4), 262–270 (2010)
 90. Kogan, F.N.: Droughts of the late 1980s in the United States as derived from NOAA polar-orbiting satellite data. *Bulletin of the American Meteorological Society* **76**(5), 655–668 (1995)
 91. Kolesov, I., Karasev, P., Tannenbaum, A., Haber, E.: Fire and smoke detection in video with optimal mass transport based optical flow and neural networks. In: 2010 IEEE International Conference on Image Processing, pp. 761–764 (2010)
 92. Kovacic, K., Ivanjko, E., Gold, H.: Computer vision systems in road vehicles: a review. In: the Proceedings of the Croatian Computer Vision Workshop (2013)
 93. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)
 94. Kumar, K.C., Bhowmick, B.: An application for driver drowsiness identification based on pupil detection using IR camera. In: Proceedings of the First International Conference on Intelligent Human Computer Interaction, pp. 73–82. Springer (2009)
 95. Lagerstrom, R., Arzhaeva, Y., Szul, P., Obst, O., Power, R., Robinson, B., Bednarz, T.: Image classification to support emergency situation awareness. *Frontiers in Robotics and AI* **3**, 54 (2016)
 96. Lai, C., Yang, J., Chen, Y.: A real time video processing based surveillance system for early fire and flood detection. In: Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE, pp. 1–6. IEEE (2007)
 97. LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E., Jackel, L.D.: Handwritten digit recognition with a back-propagation network. In: Advances in neural information processing systems, pp. 396–404 (1990)
 98. Lee, B.G., Jung, S.J., Chung, W.Y.: Real-time physiological and vision monitoring of vehicle driver for non-intrusive drowsiness detection. *IET communications* **5**(17), 2461–2469 (2011)
 99. Leira, F.S., Johansen, T.A., Fossen, T.I.: Automatic detection, classification and tracking of objects in the ocean surface from UAVs using a thermal camera. In: Aerospace Conference, 2015 IEEE, pp. 1–10. IEEE (2015)
 100. Li, L., Huang, W., Gu, I.Y., Tian, Q.: Foreground object detection from videos containing complex background. In: Proceedings of the eleventh ACM international conference on Multimedia, pp. 2–10. ACM (2003)
 101. Liao, Y.T., Huang, C.L., Hsu, S.C.: Slip and fall event detection using Bayesian belief network. *Pattern recognition* **45**(1), 24–32 (2012)
 102. Lin, C.W., Ling, Z.H.: Automatic fall incident detection in compressed video for intelligent homecare. In: Computer Communications and Networks, 2007. ICCCN 2007. Proceedings of 16th International Conference on, pp. 1172–1177. IEEE (2007)
 103. Lopez-Fuentes, L., Bagdanov, A.D., van de Weijer, J., Skinnemoen, H.: Bandwidth limited object recognition in high resolution imagery. In: Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on, pp. 1197–1205. IEEE (2017)
 104. Lu, W., Tan, Y.P.: A vision-based approach to early detection of drowning incidents in swimming pools. *IEEE transactions on circuits and systems for video technology* **14**(2), 159–178 (2004)
 105. Lu, Y.M., Do, M.N.: Multidimensional directional filter banks and surfacelets. *IEEE Transactions on Image Processing* **16**(4), 918–931 (2007)
 106. Lucas, B.D., Kanade, T., et al.: An iterative image registration technique with an application to stereo vision. *International Joint Conference on Artificial Intelligence* **81**(1), 674–679 (1981)
 107. Ma, J., Murphey, Y.L., Zhao, H.: Real time drowsiness detection based on lateral distance using wavelet transform and neural network. In: Computational Intelligence, 2015 IEEE Symposium Series on, pp. 411–418. IEEE (2015)
 108. Maalel, N., Natalizio, E., Bouabdallah, A., Roux, P., Kellil, M.: Reliability for emergency applications in internet of things. In: Distributed Computing in Sensor Systems (DCOSS), 2013 IEEE International Conference On, pp. 361–366. IEEE (2013)
 109. Maksymiv, O., Rak, T., Menshikova, O.: Deep convolutional network for detecting probable emergency situations. In: IEEE First International Conference on Data Stream Mining & Processing (DSMP), pp. 199–202. IEEE (2016)
 110. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence* **11**(7), 674–693 (1989)
 111. Martinis, S.: Automatic near real-time flood detection in high resolution x-band synthetic aperture radar satellite data using context-based classification on irregular graphs. Ph.D. thesis, lmu (2010)
 112. Mason, D.C., Davenport, I.J., Neal, J.C., Schumann, G.J.P., Bates, P.D.: Near real-time flood detection in urban and rural areas using high-resolution synthetic aperture radar images. *Geoscience and Remote Sensing, IEEE Transactions on* **50**(8), 3041–3052 (2012)
 113. Mason, D.C., Speck, R., Devereux, B., Schumann, G.J., Neal, J.C., Bates, P.D.: Flood detection in urban areas using TerraSAR-X. *Geoscience and Remote Sensing, IEEE Transactions on* **48**(2), 882–894 (2010)
 114. Medel, J.R., Savakis, A.: Anomaly detection in video using predictive convolutional long short-term memory networks. arXiv preprint arXiv:1612.00390 (2016)
 115. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 935–942. IEEE (2009)
 116. Mirmahboub, B., Samavi, S., Karimi, N., Shirani, S.: Automatic monocular system for human fall detection based on variations in silhouette area. *Biomedical Engineering, IEEE Transactions on* **60**(2), 427–436 (2013)
 117. Mobileeye: Pedestrian detection (2014). URL <http://www.mobileeye.com/technology/applications/pedestrian-detection/>
 118. Mogelose, A., Liu, D., Trivedi, M.M.: Traffic sign detection for us roads: Remaining challenges and a case for tracking. In: Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on, pp. 1394–1399. IEEE (2014)
 119. Muscio, A., Corticelli, M.A.: Land mine detection by infrared thermography: reduction of size and duration of the experiments. *IEEE Transactions on Geoscience and Remote Sensing* **42**(9), 1955–1964 (2004)
 120. Nagatani, K., Kiribayashi, S., Okada, Y., Otake, K., Yoshida, K., Tadokoro, S., Nishimura, T., Yoshida, T., Koyanagi, E., Fukushima, M., et al.: Emergency response to the nuclear accident at the Fukushima Daiichi nuclear power plants using mobile rescue robots. *Journal of Field Robotics* **30**(1), 44–63 (2013)
 121. Petak, W.J.: Emergency management: A challenge for public administration. *Public Administration Review* **45**, 3–7 (1985)

122. Piccardi, M.: Background subtraction techniques: a review. In: Systems, man and cybernetics, 2004 IEEE international conference on, vol. 4, pp. 3099–3104. IEEE (2004)
123. Piciarelli, C., Micheloni, C., Foresti, G.L.: Trajectory-based anomalous event detection. *IEEE Transactions on Circuits and Systems for Video Technology* **18**(11), 1544–1554 (2008)
124. Price, J., Maraviglia, C., Seisler, W., Williams, E., Pauli, M.: System capabilities, requirements and design of the GDL gunfire detection and location system. In: Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on, pp. 257–262. IEEE (2004)
125. Rabiner, L., Juang, B.: An introduction to hidden Markov models. *IEEE ASSP magazine* **3**(1), 4–16 (1986)
126. Radianti, J., Granmo, O.C., Bouhmala, N., Sarshar, P., Yazidi, A., Gonzalez, J.: Crowd models for emergency evacuation: A review targeting human-centered sensing. In: System Sciences (HICSS), 2013 46th Hawaii International Conference on, pp. 156–165. IEEE (2013)
127. Reisinger, K.S.: Smoke detectors: reducing deaths and injuries due to fire. *Pediatrics* **65**(4), 718–724 (1980)
128. Rinsurongkawong, S., Ekpanyapong, M., Dailey, M.N.: Fire detection for early fire alarm based on optical flow video processing. In: Electrical engineering/electronics, computer, telecommunications and information technology (ecti-con), 2012 9th international conference on, pp. 1–4. IEEE (2012)
129. Rougier, C., Auvinet, E., Rousseau, J., Mignotte, M., Meunier, J.: Fall detection from depth map video sequences. In: Toward useful services for elderly and people with disabilities, pp. 121–128. Springer (2011)
130. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Fall detection from human shape and motion history using video surveillance. In: Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on, vol. 2, pp. 875–880. IEEE (2007)
131. Rudol, P., Doherty, P.: Human body detection and geolocalization for UAV search and rescue missions using color and thermal imagery. In: Aerospace Conference, 2008 IEEE, pp. 1–8. IEEE (2008)
132. Sabokrou, M., Fayyaz, M., Fathy, M., et al.: Fully convolutional neural network for fast anomaly detection in crowded scenes. arXiv preprint arXiv:1609.00866 (2016)
133. Sakour, I., Hu, H.: Robot-assisted crowd evacuation under emergency situations: A survey. *Robotics* **6**(2), 8 (2017)
134. Saligrama, V., Chen, Z.: Video anomaly detection based on local statistical aggregates. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 2112–2119. IEEE (2012)
135. Schoeffmann, K., Hudelist, M.A., Huber, J.: Video interaction tools: A survey of recent work. *ACM Computing Surveys (CSUR)* **48**(1), 14 (2015)
136. Shah, B., Choset, H.: Survey on urban search and rescue robots. *Journal of the Robotics Society of Japan* **22**(5), 582–586 (2004)
137. Shieh, W.Y., Huang, J.C.: Falling-incident detection and throughput enhancement in a multi-camera video-surveillance system. *Medical engineering & physics* **34**(7), 954–963 (2012)
138. Siegel, R.: Land mine detection. *IEEE instrumentation & measurement magazine* **5**(4), 22–28 (2002)
139. Sixsmith, A., Johnson, N.: A smart sensor to detect the falls of the elderly. *IEEE Pervasive Computing* **3**(2), 42–47 (2004)
140. Song, Y., Njoroge, J.B., Morimoto, Y.: Drought impact assessment from monitoring the seasonality of vegetation condition using long-term time-series satellite images: a case study of Mt. Kenya region. *Environmental monitoring and assessment* **185**(5), 4117–4124 (2013)
141. Soni, B., Sowmya, A.: Classifier ensemble with incremental learning for disaster victim detection. In: Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on, pp. 446–451. IEEE (2012)
142. Soni, B., Sowmya, A.: Victim detection and localisation in an urban disaster site. In: Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on, pp. 2142–2147. IEEE (2013)
143. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. *IEEE Transactions on pattern analysis and machine intelligence* **22**(8), 747–757 (2000)
144. Suard, F., Rakotomamonjy, A., Benschraï, A., Broggi, A.: Pedestrian detection using infrared images and histograms of oriented gradients. In: 2006 IEEE Intelligent Vehicles Symposium, pp. 206–212. IEEE (2006)
145. Sultani, W., Choi, J.Y.: Abnormal traffic detection using intelligent driver model. In: Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 324–327. IEEE (2010)
146. transportation Systems, F.I.: Vehicle, pedestrian, bicycle presence sensors (2016). URL <http://www.flir.eu/traffic/display/?id=62100>
147. Thome, N., Miguet, S., Ambellouis, S.: A real-time, multiview fall detection system: A LHMM-Based approach. *IEEE Transactions on Circuits and Systems for Video Technology* **18**(11), 1522–1532 (2008)
148. Tian, H., Li, W., Ogunbona, P., Nguyen, D.T., Zhan, C.: Smoke detection in videos using non-redundant local binary pattern-based features. In: Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on, pp. 1–4. IEEE (2011)
149. Tompkin, J., Kim, K.I., Kautz, J., Theobalt, C.: Videoscapes: exploring sparse, unstructured video collections. *ACM Transactions on Graphics (TOG)* **31**(4), 68 (2012)
150. Toreyin, B.U., Dedeoglu, Y., Cetin, A.E.: HMM based falling person detection using both audio and video. In: 2006 IEEE 14th Signal Processing and Communications Applications, pp. 1–4 (2006)
151. Töreyn, B.U., Dedeoğlu, Y., Güdükbay, U., Cetin, A.E.: Computer vision based method for real-time fire and flame detection. *Pattern recognition letters* **27**(1), 49–58 (2006)
152. Trivedi, M.M., Gandhi, T., McCall, J.: Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety. *Intelligent Transportation Systems, IEEE Transactions on* **8**(1), 108–120 (2007)
153. Truong, T.X., Kim, J.M.: Fire flame detection in video sequences using multi-stage pattern recognition techniques. *Engineering Applications of Artificial Intelligence* **25**(7), 1365–1372 (2012)
154. U.S. Department of Transportation: Traffic safety facts (2015). URL <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115>
155. Utasi, Á., Benedek, C.: A multi-view annotation tool for people detection evaluation. In: Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications, p. 3. ACM (2012)
156. Van Hamme, D., Veelaert, P., Philips, W., Teelen, K.: Fire detection in color images using Markov random fields. In: International Conference on Advanced Concepts for Intelligent Vision Systems, pp. 88–97. Springer (2010)
157. Veeraraghavan, H., Masoud, O., Papanikolopoulos, N.P.: Computer vision algorithms for intersection monitoring. *Intelligent Transportation Systems, IEEE Transactions on* **4**(2), 78–89 (2003)
158. Viola, P., Jones, M.J.: Robust real-time face detection. *International journal of computer vision* **57**(2), 137–154 (2004)
159. Violence, W.H.O., Prevention, I., Organization, W.H.: Global status report on road safety 2015. World Health Organization (2015)
160. Wang, H., Finn, A., Erdinc, O., Vincitore, A.: Spatial-temporal structural and dynamics features for video fire detection. In: Applications of Computer Vision (WACV), 2013 IEEE Workshop on, pp. 513–519. IEEE (2013)
161. Wang, J.t., Chen, D.b., Chen, H.y., Yang, J.y.: On pedestrian detection and tracking in infrared videos. *Pattern Recognition Letters* **33**(6), 775–785 (2012)

162. Wang, S., He, Y., Zou, J.J., Zhou, D., Wang, J.: Early smoke detection in video using swaying and diffusion feature. *Journal of Intelligent & Fuzzy Systems* **26**(1), 267–275 (2014)
163. Wang, T., Snoussi, H.: Histograms of optical flow orientation for visual abnormal events detection. In: *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pp. 13–18. IEEE (2012)
164. Wasaki, K., Shimoi, N., Takita, Y., Kawamoto, P.N.: A smart sensing method for mine detection using time difference IR images. In: *Multisensor Fusion and Integration for Intelligent Systems, 2001. MFI 2001. International Conference on*, pp. 133–139. IEEE (2001)
165. Wojek, C., Walk, S., Schiele, B.: Multi-cue onboard pedestrian detection. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 794–801. IEEE (2009)
166. Xiong, Z., Caballero, R., Wang, H., Finn, A.M., Lelic, M.A., Peng, P.Y.: Video-based smoke detection: possibilities, techniques, and challenges. In: *IFPA, fire suppression and detection research and applications—a technical working conference (SUPDET), Orlando, FL (2007)*
167. Xu, G., Zhang, Y., Zhang, Q., Lin, G., Wang, J.: Deep domain adaptation based video smoke detection using synthetic smoke images. *arXiv preprint arXiv:1703.10729* (2017)
168. Xu, Z., Xu, J.: Automatic fire smoke detection based on image visual features. In: *Computational Intelligence and Security Workshops, 2007. CISW 2007. International Conference on*, pp. 316–319. IEEE (2007)
169. Xu, Z., Xu, J.: Automatic fire smoke detection based on image visual features. In: *Computational Intelligence and Security Workshops, 2007. CISW 2007. International Conference on*, pp. 316–319 (2007)
170. Yamato, J., Ohya, J., Ishii, K.: Recognizing human action in time-sequential images using hidden Markov model. In: *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, pp. 379–385 (1992). DOI 10.1109/CVPR.1992.223161
171. Yang, L., Yang, S.H., Plotnick, L.: How the internet of things technology enhances emergency response operations. *Technological Forecasting and Social Change* **80**(9), 1854–1867 (2013)
172. Ye, W., Zhao, J., Wang, S., Wang, Y., Zhang, D., Yuan, Z.: Dynamic texture based smoke detection using surfacelet transform and hmt model. *Fire Safety Journal* **73**, 91–101 (2015)
173. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. *Acm computing surveys (CSUR)* **38**(4), 13 (2006)
174. Yu, C., Mei, Z., Zhang, X.: A real-time video fire flame and smoke detection algorithm. *Procedia Engineering* **62**, 891–898 (2013)
175. Yuan, F.: Video-based smoke detection with histogram sequence of LBP and LBPV pyramids. *Fire Safety Journal* **46**(3), 132–139 (2011)
176. Yun, K., Jeong, H., Yi, K.M., Kim, S.W., Choi, J.Y.: Motion interaction field for accident detection in traffic surveillance video. In: *2014 22nd International Conference on Pattern Recognition (ICPR)*, pp. 3062–3067. IEEE (2014)
177. Zadeh, L.A.: Fuzzy sets. *Information and Control* **8**(3), 338–353 (1965)
178. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics SMC-3*(1), 28–44 (1973)
179. Zaklouta, F., Stanculescu, B.: Warning traffic sign recognition using a HOG-based Kd tree. In: *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pp. 1019–1024. IEEE (2011)
180. Zecha, D., Greif, T., Lienhart, R.: Swimmer detection and pose estimation for continuous stroke-rate determination. In: *IS&T/SPIE Electronic Imaging*, pp. 830,410–830,410. International Society for Optics and Photonics (2012)
181. Zhang, Q., Xu, J., Xu, L., Guo, H.: Deep convolutional neural networks for forest fire detection. In: *Proceedings of the 2016 International Forum on Management, Education and Information Technology Application*. Atlantis Press (2016)
182. Zhang, T., Wang, J., Xu, L., Liu, P.: Fall detection by wearable sensor and one-class SVM algorithm. In: *Intelligent computing in signal processing and pattern recognition*, pp. 858–863. Springer (2006)
183. Zhou, S., Shen, W., Zeng, D., Fang, M., Wei, Y., Zhang, Z.: Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Processing: Image Communication* **47**, 358–368 (2016)